Spring 2024

Evaluating the Influence of Operator Practices on Hydraulic Shovel Productivity and Energy Consumption using Telemetry Data

Noah Adekunle Aluko
Missouri University of Science and Technology

EVALUATING THE INFLUENCE OF OPERATOR PRACTICES ON HYDRAULIC SHOVEL PRODUCTIVITY AND ENERGY CONSUMPTION USING TELEMETRY DATA

by

NOAH ADEKUNLE ALUKO

A THESIS

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE IN MINING ENGINEERING

2023

Approved by:

Kwame Awuah-Offei, Advisor
Akim Adekpedjou
Samuel Frimpong

ABSTRACT

Hydraulic shovels are increasingly being adopted in the mining industry, which is known for its significant energy consumption and carbon emissions. Technological improvements in hydraulic shovels aim at bolstering energy efficiency and productivity because of the drive to curb energy consumption and enhance energy efficiency. However, one aspect that is often overlooked is the vital role of operators. Although the influence of operators' practices on excavators' energy efficiency has been acknowledged in past studies, there is a paucity of quantitative research assessing this influence in the context of hydraulic shovels.

The main objectives of this study are to: 1) develop algorithms for meaningful data extraction from shovel telemetry; and 2) test the hypothesis that operator practices significantly influence hydraulic shovel energy efficiency, focusing on identifying key differentiating parameters. This study collected telemetry data from a 40.5 yd^3 bucket hydraulic shovel and developed several algorithms to extract meaningful data for comprehensive statistical analysis to fulfill the first objective. This study utilized statistical tests, including equality of means, to determine differences in operators' energy efficiencies and further employed statistical data analysis techniques such as correlation and difference regression analysis to identify key parameters influencing these variations.

The analysis concluded that payload is the most significant variable influencing the differences in energy efficiencies of operators, as it consistently appears in the comparison across all operators, while variables such as boom angle, swing-out time, and digging time were less consistent explaining differences in operators' energy efficiency.

ACKNOWLEDGMENTS

I am profoundly grateful to God for being my sanctuary in times of difficulty, my guiding light when I felt lost, and my anchor throughout my academic endeavors. I am immensely thankful to my esteemed advisor, Dr. Kwame Awuah-Offei, for his unwavering support, insightful advice, and constant encouragement throughout my academic journey. His wisdom and guidance have significantly influenced this thesis and shaped my academic path. I am equally appreciative of Drs. Akim Adekpedjou and Samuel Frimpong, my committee members, for their erudite comments and constructive criticism, which have enhanced the quality and rigor of my work.

I sincerely thank ESCO WEIR and the Mining and Explosives Department at the Missouri University of Science and Technology for their generous provision of essential resources, without which this study would not have been possible. I am also grateful to my friends - Salami Oluwafemi, Oluwasegun Agagu, Hussam Altalhi, and Theophilus Mensah - for their unwavering positivity, steadfast support, and constant encouragement. Your presence has been instrumental in helping me overcome the challenges I encountered during this academic journey.

Finally, I am deeply indebted to my family for their unconditional love, prayers, and support. Their faith in my capabilities has motivated me to ascend heights I once perceived as insurmountable. As I stand at the culmination of this journey, looking back at the path trodden, I am reminded that I am but the result of many hands, each contribution invaluable. This thesis is not solely a testament to my perseverance but, more importantly, to the collective effort of those who believed in my dream.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

NOMENCLATURE

Symbol	Description
η	Energy Efficiency
η_l	Energy per Unit Loading Rate
$\eta_l(i)$	Energy per Unit Loading Rate of Cycle *i*
$\Delta\eta_l$	Energy per Unit Loading Rate Difference Between Operators
P	Payload or Amount of Materials Loaded
$P(i)$	Payload of Cycle *i*
E_t	Total Energy Expended
$E_t(i)$	Total Energy Expended in Cycle *i*
E_b	Energy Expended by the Boom
$E_b(i)$	Energy Expended by the Boom in Cycle *i*
Γ_i	Torque Produced by the Boom in Cycle *i*
θ_i	Angular Displacement of the Boom in Cycle *i*
t_c	Cycle Time
$t_c(i)$	Cycle *i* Time
σ_n	Standard Deviation of Energy per Unit Loading Rate
N_{cop}	Number of Cycles of Operators
${}^nC_k$	Number of Combinations
$!$	Factorial
$k_{0\text{-}9}$	Coefficients of Regression
ΔP	Payload Difference Between Operators

ΔC_t Cycle Time Difference Between Operators

ΔC_t Cycle Time Difference Between Operators

ΔSi_t Swing-in Time Difference Between Operators

ΔDi_t Digging Time Difference Between Operators

ΔSo_t Swing-out Time Difference Between Operators

ΔD_h Dump Height Difference Between Operators

ΔB_a Boom Angle Difference Between Operators

ΔSt_a Stick Angle Difference Between Operators

ΔSi_a Swing-in Angle Difference Between Operators

1. INTRODUCTION

1.1. BACKGROUND

The efficient operation of hydraulic shovels is crucial for the productivity and profitability of mining, construction, and earthmoving industries, as they play a vital role in extraction and loading processes. Hydraulic excavators (in either shovel or backhoe configurations) have significant importance in the mining industry due to their efficiency, versatility, power, design for heavy-duty excavation, and material handling in challenging environments (Caterpillar, 2023; Hitachi, 2023; Komatsu, 2023). Over the years, hydraulic shovels have evolved with technological advancements, leading to improved efficiency, increased payload capacity, and reduced energy consumption. However, one often-overlooked aspect that affects the performance of these machines is the influence of operators' practices on energy consumption and productivity (Komljenovic et al., 2010).

Operators play a crucial role in controlling and maneuvering these machines. Consequently, their skills and practices directly affect energy consumption and productivity (Dindarloo et al., 2016). In addition, different operators may exhibit varying levels of experience, knowledge, and ability to adapt to different working conditions, which can significantly impact the efficiency of shovels (Awuah-Offei & Frimpong, 2007). Therefore, understanding and evaluating these factors is critical for optimizing the use of these machines, reducing operating costs, and minimizing environmental impacts.

The mining industry is a significant energy consumer globally. In the United States, to be precise, the mining industry consumes approximately 1.4% of the total energy produced (U.S. Energy Information Administration, 2021). With the increasing need for more sustainable energy practices, the mining industry constantly seeks to reduce energy consumption and improve energy efficiency. Optimizing the energy consumption of hydraulic shovels can contribute to significant energy savings, thereby reducing the overall environmental footprint of mining operations.

Recent advancements in telemetry technology have made it possible to collect real-time data on hydraulic shovel performance. These advances include advances in edge computing,[1] the Internet of Things (IoT), wireless communications technology, sensor technology, and machine learning (or artificial intelligence) (Aguirre-Jofré et al., 2021; Jacobs et al., 2018; Komatsu, 2023; Mining Technology, 2022). Modern telemetry systems can measure various parameters, such as fuel consumption, engine load, and cycle times, to provide insights into the machine's energy efficiency and productivity. Examples of the monitoring systems in the industry using telemetry include Caterpillar's vital information management systems (VIMS), Komatsu's LINCS II and monitoring system, Hitachi's MIC, and Motion metrics ShovelMetrics™. In addition, the data collected from telemetry systems could be analyzed to evaluate the influence of the practices of operators on the machine's performance.

In summary, analyzing the influence of the operators' practices on hydraulic shovel energy consumption and productivity constitutes an essential field of study with significant implications for the competitiveness and sustainability of industries relying on

---

[1] Edge computing involves local data processing on the telemetry device, resulting in improved response time and reduced data transmission to a central server.

these machines. The use of telemetry, combined with advanced statistics, data analytics, and machine learning techniques, presents a promising approach to understanding and optimizing the part of operators' practices in the energy efficiency (performance) of hydraulic shovels.

1.2. PROBLEM STATEMENT

Mining demands a considerable amount of energy and accounts for a significant fraction of the world's energy consumption (Igogo et al., 2021). In the United States, mining is the second-largest portion of annual industrial energy consumption within the industrial sector, contributing to approximately 9% of the total energy expenditure (U.S. Energy Information Administration, 2022). As a result, the mining industry is responsible for a considerable portion of global carbon emissions, contributing to climate change and environmental degradation (Norgate & Haque, 2010). However, investing in "state-of-the-art equipment and conducting further research" could significantly diminish energy usage to roughly 47% of current levels (U.S. Department of Energy, 2007). The analysis suggests that implementing best practices could result in a total energy savings potential of 667 Trillion Btu/yr. Specifically, adopting best practices could save 258 Trillion Btu/yr (U.S. Department of Energy, 2007). In contrast, research and development efforts to improve mining technologies could save an additional 409 Trillion Btu/yr (U.S. Department of Energy, 2007). Furthermore, these practical energy savings could reduce approximately 40.6 million tonnes of CO_2 emissions (U.S. Department of Energy, 2007).

The two categories of mining machinery that present the most significant prospect for energy conservation in mining are grinding and material handling machinery (U.S.

Department of Energy, 2007). Adopting best practices and integrating novel developments via research and development could lead to an energy saving of 111 trillion Btu per year in materials handling operations and 356 trillion Btu per year in grinding. In addition, By minimizing the energy usage of these two procedures to their practical lowest, the mining sector could conserve approximately 37% of its present energy expenditure (U.S. Department of Energy, 2007). Figure 1.1 shows the energy conservation prospects for the mining sector's top 10 energy-consuming processes in the U.S.

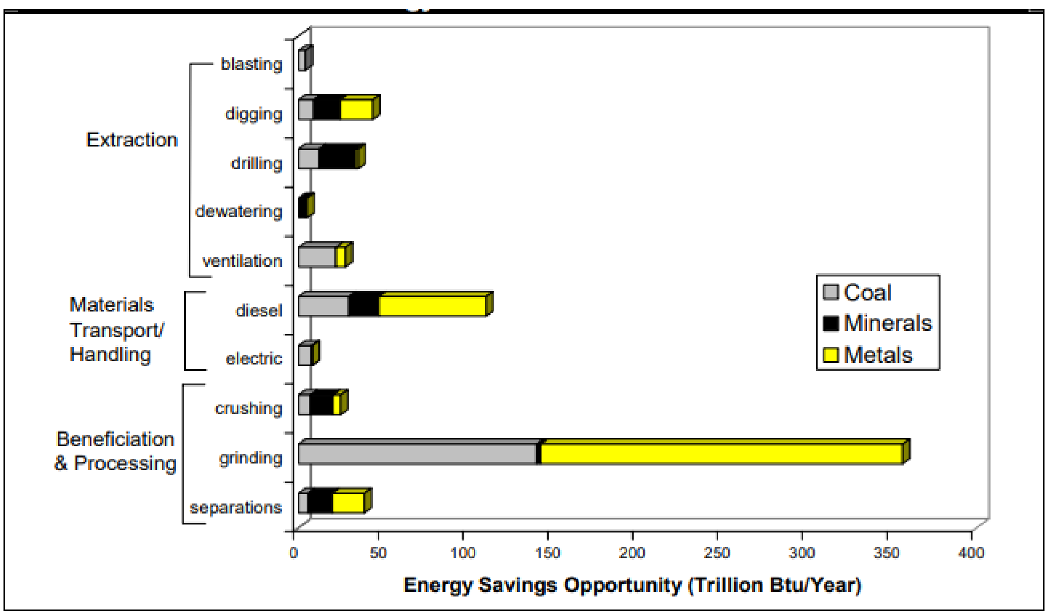

Figure 1.1 Opportunity for energy conservation in the U.S. mining sector, adapted from (U.S. Department of Energy, 2007)

Table 1.1 shows that the hydraulic shovel is the second most significant energy-consuming equipment in surface mines, as estimated by the SHERPA Mine Cost Estimating Model. Given the pressing demand to lower CO_2 emissions, the substantial capital investment required, the escalating energy costs, and the significant impact on mining profitability, any enhancements in productivity and energy efficiency will bring enormous advantages to mining operations (Abdi-Oskouei, 2013).

Table 1.1 Material handling equipment assumed (U.S. Department of Energy, 2007)

Equipment	Units	Btu per hr (single unit)	Daily hour per unit
Front-end Loaders	5	3,640,682	14
Water Tankers	1	1,502,182	2.94
Rear Dump Trucks	11	1,656,897	14
Hydraulic Shovel	1	4,102,318	9.38
Graders	1	618,841	0.56
Bulldozer	2	5,115,421	14
Service Trucks	2	339,364	14
Pick-up Trucks	8	207,112	14

The energy efficiency of loading equipment in mining greatly relies on the practices of the operators. (Abdi-Oskouei, 2013; Awuah-Offei, 2016; Komljenovic et al., 2010; Lumley & Lumley, 2005; Patnayak et al., 2008). The productivity of mining shovels can vary significantly depending on the operator, even under the same operating conditions (Awuah-Offei, 2016; Patnayak & Tannant, 2005; Vukotic, 2013; Yaghini et al., 2020). It is possible to improve energy efficiency and reduce cost by clearly understanding the interaction between energy efficiency and the practices of operators (Abdi-Oskouei, 2013). However, there is still a scarcity of research dedicated to numerically examining the influence of operators’ practices on the energy efficiency of hydraulic shovels and the causes behind disparities. Past studies have highlighted the substantial impact of operator skills and practices on electric rope shovel and dragline productivity (Abdi-Oskouei & Awuah-Offei, 2016; Awuah-Offei & Frimpong, 2007). However, the kinematics and dynamics of electric rope shovels and draglines differ from hydraulic excavators and will likely mean different actuators controlled by the operator will determine any effects on energy efficiency and productivity. Therefore, this study uses statistical tools to investigate the connection between the energy efficiency of hydraulic shovels and operator practices. The ultimate goal is to devise a technique capable of assessing the influence of operators' practices on the energy efficiency of hydraulic shovels.

1.3. OBJECTIVES AND SCOPE OF THIS RESEARCH

The primary goal of this work is to evaluate the influence of operators' practices on hydraulic shovel energy consumption and productivity. The precise objectives of the research are to:

1. Develop algorithms to extract meaningful data (i.e., Data per cycle useful for understanding operator effects on energy efficiency and productivity) out of the shovel data for statistical data analysis; and
2. Examine the hypothesis that the abilities and practices of hydraulic shovel operators influence the shovel's energy efficiency. Hence, identify critical factors responsible for the disparities in operators' energy efficiencies.

In this study, the researcher delves deeper into investigating the influence of operators' practices on hydraulic shovel energy consumption and productivity using data collected by telemetry. The author explores the methods and techniques employed to collect and analyze data, discusses the connection between the practices of operators and machine performance, and identifies key parameters that differentiate energy efficiency. Ultimately, this analysis aims to contribute to the ongoing efforts to improve energy efficiency and sustainability in the mining and industries.

Abdi-Oskouei (2013) did a similar study on dragline energy efficiency. However, to this author's knowledge, no similar study has been done on hydraulic shovels to pinpoint the key parameters causing the difference in the energy efficiency of operator. This work improves upon the approach that Abdi-Oskouei (2013) proposed, using theoretically sound approaches to linear-circular regression that Abdi-Oskouei (2013) ignored.

In essence, to accomplish the first objective of this research, this researcher developed algorithms as MATLAB functions to extract relevant data necessary for the analysis and improved upon the difference linear regression approach proposed by Abdi-Oskouei (2013) by using theoretically sound approaches to linear-circular regression to achieve the second objective. The tests and studies are all carried out on the field dataset obtained from the same hydraulic shovel.

1.4. STRUCTURE OF THE THESIS

This thesis consists of six sections. Section 1 presents the introduction, which includes the background, problem statement, objectives and scope of this study, and structure of the thesis. Section 2, which is the literature review, presents a review of past relevant studies such as hydraulic shovel operation, the general application of telemetry in mining, telemetry focused on excavators, the significance of energy efficiency, factors affecting shovel energy efficiency, and operator practice effects on mining equipment performance. Information about sensor data structure, cycle identification, sampling, validation, and extraction of meaningful data is presented in Section 3. Section 4 provides an initial analysis of the sensor data employed in the case studies. This analysis encompasses the examination of summary statistics and the identification and subsequent replacement of outliers within the dataset. Section 5 presents the influence of operators' practices on hydraulic shovels' energy efficiency. This section further presents a procedure, which includes a study case for assessing key parameters that account for the observed differences in energy efficiencies of the shovel operators. Finally, Section 6 concludes this research and provides recommendations for future study.

2. LITERATURE REVIEW

This section of the thesis reviews the relevant literature to serve as the basis for the research. The review covers telemetry and its applications in shovel operations, the basics of hydraulic shovel operations, energy efficiency, in general and specific to hydraulic shovels, and the influence of the practices of operators on shovel (energy) efficiency. The review underscores the existing gaps in the literature, thereby placing this research in the context of the literature.

2.1. GENERAL TELEMETRY IN MINING

The term “telemetry” commonly refers to the transfer of data through wireless means like radio, ultrasonic, or infrared systems from remote or inaccessible locations to a central system for analysis and monitoring (Chetty, 1982; Srivastava, 1975). Nonetheless, it includes data transmitted via other channels such as telephonic or computer networks, wired communication systems like power line carriers, or optical connections. Furthermore, telemetry involves the utilization of a physical instrument known as a “telemeter” that incorporates a sensor, a transmission route, and a device for display, recording, or control.(Bakshi & Bakshi, 2020). In addition, telemetry often employs electronic devices, which may be wireless, hardwired, analog, or digital (Figure 2.1). Other mechanical, hydraulic, and optical technologies are also used in telemetry (Bakshi & Bakshi, 2020).

Initially, telemetry in mines was designed for a dual purpose, including monitoring physical parameters and preventing potential hazards (Chetty, 1982).

However, in the mining industry, telemetry systems have become increasingly crucial, offering a range of benefits, from improved safety to enhanced productivity and reduced operational costs (Kiziroglou et al., 2017). As a result, it has evolved significantly over the past few decades. In recent years, the mining industry has seen a significant increase in the application of telemetry systems, which is driven by advances in communication technologies and the need for improved safety, productivity, and environmental monitoring (Duarte et al., 2022; F. Sánchez & Hartlieb, 2020a). Major applications of telemetry in mining include:

- **Safety**. Telemetry enables real-time monitoring of equipment and personnel, reducing the risk of accidents and injuries by detecting unsafe conditions and providing early warnings (Akkaş, 2018; Sadeghi et al., 2022; Zhu & You, 2019).
- **Productivity**. Telemetry systems can optimize mining processes and equipment utilization, increasing productivity and reducing operational costs (Kiziroglou et al., 2017; McKinnon, 2022; McKinsey, 2018; Peterson, 1986).
- **Environmental Monitoring**. Telemetry systems can monitor environmental parameters and help mitigate potential environmental impacts, such as air and water pollution, through early detection and response (Jha & Tukkaraja, 2020; Jo & Khan, 2018; Minhas et al., 2018).

Several studies have examined the use of electronics and sensing technology in mining operations (Chetty, 1982; Dong et al., 2017; Fantini et al., 2017; Kalinowski et al., 2022; Lanciano & Salvini, 2020; Moridi et al., 2015; Peterson, 1986; Ranjan et al., 2020; Srivastava, 1975; Yuval et al., 2019). For example, Ruff & Hession (2001) outlines experiments conducted on commercially available Radio Frequency Identification

(RFID) systems, followed by developing a tailored RFID system suitable for surface and underground mining equipment applications. Similarly, Nguyen et al. (2020) employed vibration sensors to gauge ground vibrations induced by blasts.

Figure 2.1 Telemetry system overview (Sierra Wireless, 2015)

Advancements in sensing technologies, communication, and analytics have significantly improved telemetry applications in the mining industry. Furthermore, integrating Internet of Things (IoT) devices, advanced machine learning algorithms, and cloud-based data storage has significantly improved the functionality of telemetry systems. As a result, these systems make it possible for real-time analysis and decision-making. Hence, they facilitate the development of predictive models for equipment maintenance, process optimization, and hazard identification (Liu et al., 2018).

While the vast opportunities offered by this field of research are undeniable, it is crucial to acknowledge the immense volume of data generated by these technologies. Therefore, addressing issues related to data storage and, more importantly, data management is paramount. Solutions may lie in the capabilities of artificial intelligence and machine learning (Duarte et al., 2022; Mansouri et al., 2020; Pishgar et al., 2021).

Although telemetry in mining has made considerable progress, there remain areas for further research and development. These opportunities include creating standardized protocols and frameworks to facilitate interoperability among various telemetry systems and mining equipment, ensuring seamless integration; examining and addressing cybersecurity risks linked to the growing connectivity and dependence on telemetry systems in mining operations; and investigating energy-efficient telemetry solutions, particularly for remote and off-grid mining sites. Also, research that extracts actionable decisions from telemetry data beyond what the systems were designed to address can extend telemetry systems' usefulness and benefits relative to costs. Pursuing these research avenues could significantly augment the advantages of telemetry in the mining industry.

2.2. TELEMETRY FOCUSED ON EXCAVATORS

The application of telemetry and sensing technologies in the mining industry has seen substantial growth in recent years, and this is particularly evident in their use on excavators. Recent developments in mining include new sensing and monitoring systems that enhance equipment performance (F. Sánchez & Hartlieb, 2020b). Early telemetry systems involved simple data-logging devices that collected information on machine performance, which had to be manually retrieved and analyzed.

With advancements in sensing technologies and data communication, modern telemetry systems have become more sophisticated, enabling real-time monitoring and control of mining equipment. Examples of such technologies include Caterpillar's vital information management systems (VIMS), Komatsu's LINCS II and monitoring system, Hitachi's MIC, Drives & Controls Service AccuWeigh™, and Motion metrics ShovelMetrics™. These sensing and monitoring technologies allow mine operators to track the location of equipment, ensure its proper functioning, and make informed decisions regarding equipment utilization and hiring. In addition, by optimizing equipment utilization and preventing redundant hires, telemetry helps reduce equipment hire costs and improve overall operational efficiency.

Excavators are commonly used for material handling in unstructured environments like mining and construction. However, operating them in these practical settings can pose challenges due to severe conditions like rock fall, soil subsidence, and excessive particulate matter, resulting in potential fatalities and injuries (Zhang et al., 2021). Therefore, telemetry has been integrated into mining operations to ensure safe, efficient, and sustainable operations. Advancements in sensing and monitoring technologies drive this integration. For example, Kiziroglou et al. (2017) noted that real-time monitoring of mining operations, mines can improve productivity and reduce operational costs. Similarly, Pan (2005) suggests that condition monitoring through telemetry can enhance the maintenance of heavy machinery, including excavators, and can help in predicting potential failures, thus reducing downtimes. As a result, Original

Equipment Manufacturers (OEMs) increasingly embed telemetry sensors into their products and retrofit existing equipment with telemetry devices. However, the challenge lies in accessing telematics data from multiple sources to facilitate effective decisions, as each OEM often has its own proprietary technology.

Telemetry and sensing technologies on excavators are a subset of this broader integration. Excavators in mining are subject to harsh conditions, leading to potential equipment failure. Thus, telemetry is invaluable for predictive maintenance. For example, workers at the mine were usually required to go inside the primary crusher to retrieve a lost tooth[2]. However, with the current integration of monitoring sensors, workers can be kept out of harm's way while ensuring uninterrupted operations (Motion Metrics, 2023). In addition, real-time monitoring and analysis of operational parameters of excavators, such as engine temperature, hydraulic pressure, vibrations, and energy consumption, can help identify irregular patterns that may indicate possible equipment failure (Chen et al., 2010; Lazarević et al., 2016; Pan, 2005; X. Zhou & Lei, 2021).

Telemetry systems in excavators are unique only in the type of sensors used for data acquisition[3]. The commonly used telemetry systems by draglines and cable shovels in the mining industry are the AccuWeigh™ and ShovelMetrics™ monitoring systems. These monitoring systems comprise multi-sensors that provide comprehensive real-time performance data to the operator, enabling informed decision-making in mining

---

[2] Tooth missing or broken from the bucket of the excavator during digging or loading process. In mining, the tooth is commonly referred to as a component of ground engagement tool (GET).

[3] The communication and data storage systems are similar to other mine telemetry systems.

operations. The Accuweigh™ monitoring system is a microprocessor-based remote observation tool that provides continuous performance data in real time. Its algorithms enable monitoring of the primary machine[4] signals, which include motor current and voltage, and convert these signals into meaningful information. This information includes the exact location of the bucket, payload, and swing angle displayed to the operator on a digital screen. It also monitors machine overloading and alerts the operator in real time.

Similarly, the ShovelMetrics™, equipped with a high-resolution 3D bucket camera, pressure sensors, and accelerometers, provides precise data about the bucket's position in real-time. This feature, coupled with its AI-enhanced fragmentation analysis, informs personnel about the size of the material in the bucket and allows them to make adjustments on the spot to improve efficiency. Additionally, with the system's ability to detect worn and broken ground engagement tool (GET) components, operators are aware of the need for maintenance before the bucket's performance diminishes. In essence, both systems provide real-time feedback and critical data points on key performance indicators, which include payload, swing angles, energy components, cycle time components, and potential operational issues. Thus, allowing for proactive decision-making, increased efficiency, and enhanced safety in mining operations.

In summary, telemetry systems on excavators offer numerous benefits, such as improved safety, enhanced efficiency, and cost savings. These systems enable real-time monitoring of equipment and environmental conditions, allowing for proactive measures to prevent accidents and optimize operations. In addition, by leveraging telemetry data

---

[4] The machine referred to here can either be a cable shovel or a dragline.

and analytics, mine sites can make better equipment hiring decisions, reduce costs, and improve overall performance.

2.3. BASICS OF HYDRAULIC SHOVEL OPERATION

Hydraulic shovels, initially utilized in the construction industries and now more commonly found in surface mining operations across the globe, have become an integral part of both sectors, offering high-capacity and efficient excavation capabilities (Patnayak, 2006). In addition, the versatility and maneuverability of hydraulic shovels have led to significant advancements in these sectors, making them essential pieces of equipment. Hydraulic shovels are earthmoving machines that consist of a boom, stick, bucket, hydraulic cylinders, and an undercarriage for mobility. Figure 2.2 illustrates the nomenclature and assemblies of hydraulic shovels.

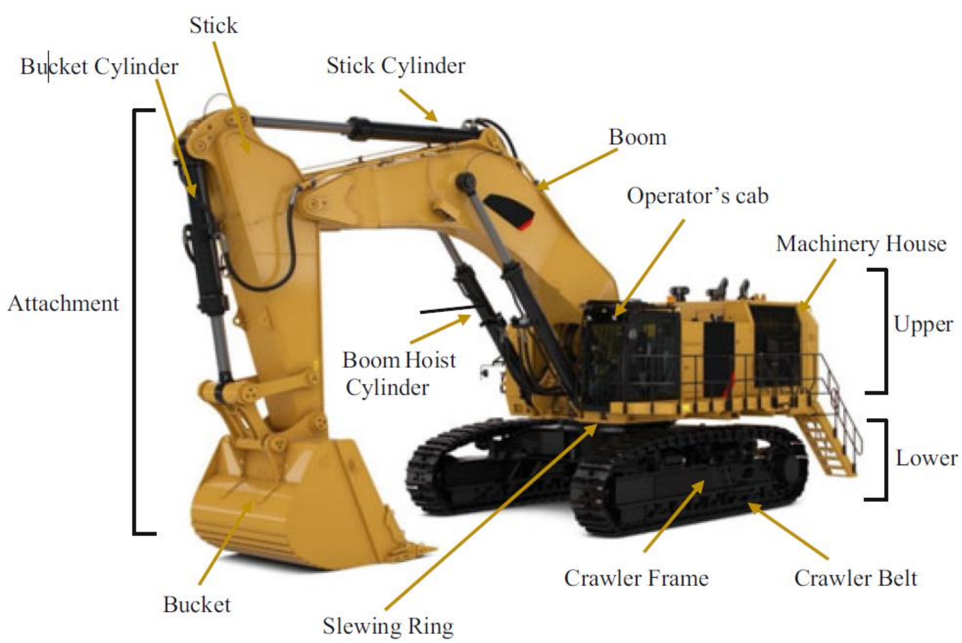

Figure 2.2 Nomenclature of a hydraulic shovel (Soofastaei et al., 2018)

These large excavators operate on hydraulics principles, using pressurized fluid to transmit power and control motion. In hydraulic shovels, the entire digging process is controlled by hydraulic systems driven by diesel engines. They come in backhoe and front shovel configurations. In contrast to their predecessor, the cable shovel, hydraulic shovels offer improved precision, flexibility, control, and efficiency (Andreev et al., 2017; Patnayak, 2006). The design of hydraulic shovels focuses on maximizing efficiency, durability, and ease of maintenance (Caterpillar, 2023; Hitachi, 2023). Material selection and structural analysis are crucial for ensuring the reliability and longevity of the components (Caterpillar, 2023). Efficient hydraulic shovel operation relies on proper operating techniques, such as bucket positioning and digging strategies.

Besides the forces used to move the machine during operation, forces acting on the shovel bucket are the most important forces dictating energy consumption. The literature shows that six unique forces act on the bucket during digging, as Hemami et al. (1994) indicated. Figure 2.3 illustrates these forces. The first force, f_1, involves the exertion required to counteract the weight of the material loaded into the bucket. This force significantly contributes to the total digging resistance. Another force, f_2, arises due to the material's resistance to compression by the base of the shovel bucket. However, Hemami et al. (1994) and Takahashi et al. (1998) propose this force to be negligible, as the bucket base does not significantly compress the material (soil or muck pile) in most operational trajectories. The third force, f_3, is attributed to the friction between the material and the bucket as the material slides down the bucket's inner surface. This force is often considered a component of the overall cutting resistance. The fourth force, f_4, results from the resistance encountered at the bucket tip during soil cutting or muck pile penetration. Research identifies this force, with f_1, as essential in determining digging resistance (Hemami et al., 1994; Takahashi et al., 1998). The fifth force, f_5, corresponds to the inertial force of the loaded material within the bucket. This force is often considered inconsequential if no significant acceleration occurs during excavation. The final force, f_6, relates to the energy necessary to mobilize the empty bucket. This force can be accurately determined once the bucket's weight is established.

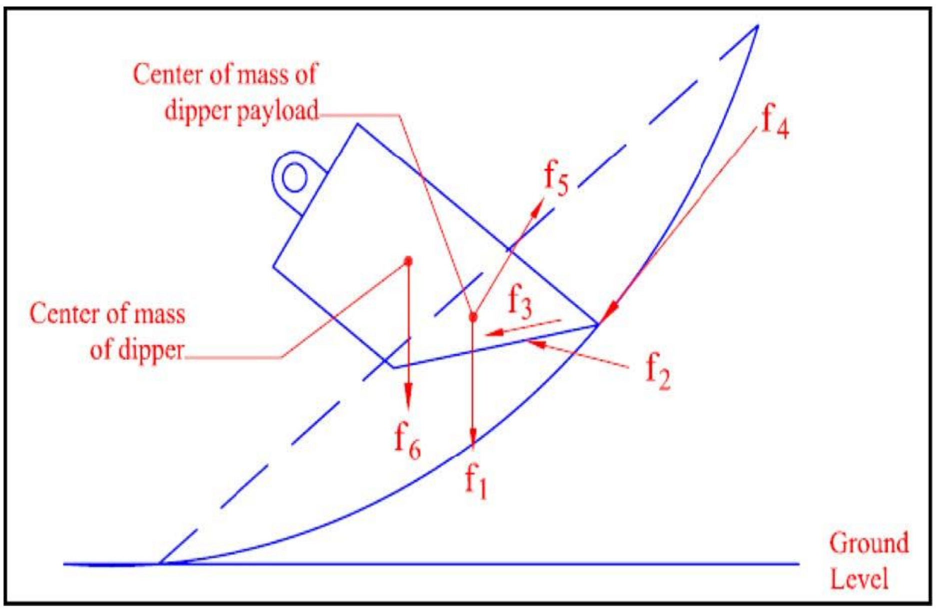

Figure 2.3 Forces acting on the bucket of the shovel during excavation (Awuah-Offei et al., 2009)

The standard operation of a shovel, excluding the traversing, consists of a cyclic process (Figure 2.4). The complete cycle operation of a hydraulic shovel involves swinging in and positioning the bucket, filling the empty bucket with materials by digging the face, swinging out, and dumping the materials to commence the next cycle. Hydraulic shovels typically have a capacity ranging from 7 yd^3 (5 m^3) to 70 yd^3 (54 m^3)

for standard rock applications (Caterpillar, 2020). Due to its large size and high production rate, this machine is a significant energy consumer in mining operations. For example, a 14.4 yd^3 bucket capacity shovel can consume 26-40 gal/hr of fuel (Awuah-Offei et al., 2011).

Factors affecting hydraulic shovel performance include bucket capacity (payload), cycle time, energy efficiency, and machine availability (Dindarloo et al., 2016; Soofastaei et al., 2018). The techniques for improving performance include payload monitoring, machine condition monitoring, and operator training (Aguirre-Jofré et al., 2021; Chen et al., 2010). In addition, simulation and modeling tools, such as discrete event simulation and finite element analysis, can aid performance analysis and optimization (Awuah-Offei et al., 2011; Raj et al., 2009). Case studies on performance improvement demonstrate the potential benefits of implementing these techniques.

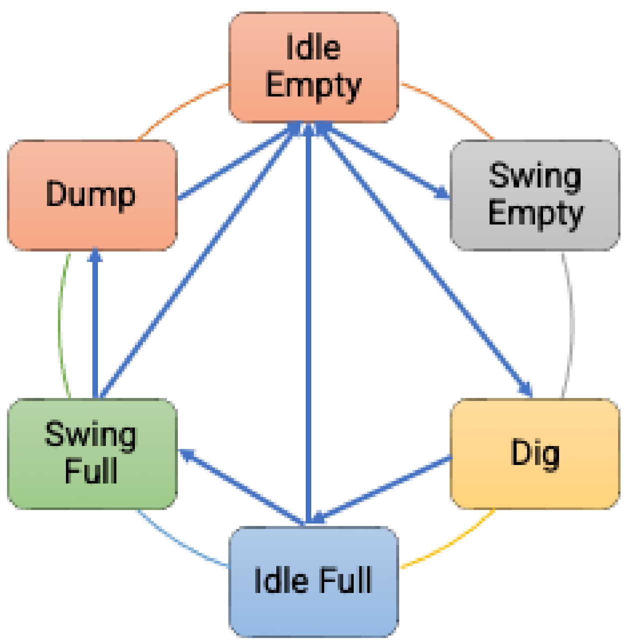

Figure 2.4 Hydraulic shovel sequence of operation

2.4. ENERGY EFFICIENCY IN MINING

The mining industry is essential for extracting minerals and materials to sustain the global economy. However, the industry significantly contributes to global energy consumption, accounting for 15% of worldwide electricity consumption, 38% of worldwide industrial energy utilization, and 11% of worldwide energy expenditure (Igogo et al., 2021). The high energy consumption contributes to greenhouse gas emissions, operational costs, and environmental impacts. Energy costs form a substantial part of the mining industry’s budget, accounting for an average of 15% to 40% of the total operating expenses (Igogo et al., 2021; Maennling & Toledano, 2018). This considerable investment in energy production, primarily relying on fossil fuels, leaves the mining sector extremely vulnerable to fluctuations in the fossil fuel market (Igogo et al., 2021). Projections anticipate a potential growth of up to 36% in the energy demand for mining activities by 2035 (Igogo et al., 2021; Maennling & Toledano, 2018).

Therefore, improving energy efficiency in mining has become a priority for stakeholders, including governments, mining companies, and environmental groups. With growing concerns about climate change, resource depletion, and rising energy costs, it is becoming increasingly critical for the mining industry to enhance its energy efficiency. Enhancing energy efficiency in mining can help reduce operational costs, mitigate environmental impacts, and contribute to global sustainability goals.

Mining energy consumption can be categorized into four main areas: extraction, materials handling, comminution, and mine support services (Awuah-Offei, 2018b), with comminution and material handling being the two major energy-intensive processes in mining operations. Comminution, which includes crushing, grinding, and separation, is among the highest energy-consuming mining processes, accounting for up to 50% of total energy consumption (Soofastaei & Fouladgar, 2022a). On the other hand, materials handling, which includes loading and haulage of ore and waste materials, accounts for about 10-20% of the total energy (Kecojevic et al., 2014; Soofastaei & Fouladgar, 2022a). The mining sector traditionally relies on diesel, natural gas, and grid electricity for its energy needs (Igogo et al., 2021; Soofastaei & Fouladgar, 2022a).

However, there has been a notable shift towards natural gas and grid electricity (Soofastaei & Fouladgar, 2022a). In addition, the industry is witnessing rising adoption of renewable energy sources to fuel its operations (Igogo et al., 2021; Kirk & Cannon, 2020). This significant transition is influenced by various factors, such as the substantial energy costs associated with mining, the industry's commitment to achieving environmental objectives, and the societal considerations related to mining operations (Igogo et al., 2021). Furthermore, this evolution in energy use underlines the mining industry's efforts to balance operational demands with sustainability and social responsibilities (Igogo et al., 2021).

As far as the mining industry is concerned, energy efficiency is typically gauged by the ratio of valuable work accomplished to the input of energy (Awuah-Offei, 2016). In the mining industry, the volume of output, such as rock tonnage or metal grammage, often serves as a surrogate for valuable work done (Abdi-Oskouei & Awuah-Offei, 2014; Awuah-Offei, 2016; Odhams et al., 2010). In addition, the amount of diesel consumed by haulage trucks is frequently used as an energy input indicator (Awuah-Offei, 2016; Awuah-Offei et al., 2011; Motlogelwa & Minnitt, 2013). Various factors such as ore grade, depth, mine layout, extraction technologies, and equipment efficiency influence energy consumption in mining (Koppelaar & Koppelaar, 2016; Soofastaei & Fouladgar, 2022a). For example, extracting lower ore grades would require more energy (Igogo et al., 2021; Lezak et al., 2019; Norgate & Haque, 2010). At the same time, deeper mines necessitate more energy for ventilation and materials handling (Awuah-Offei, 2018b; Demirel, 2018; Koppelaar & Koppelaar, 2016; Levesque et al., 2014). Mine layout can also impact energy efficiency, as longer haul distances and more complex road networks increase energy requirements.

Efforts to boost energy efficiency in the mining sector address every facet, ranging from waste heat recovery and electricity demand management to the production of energy from residues (waste products), mine drainage reduction, and ventilation improvement (Awuah-Offei, 2016; Levesque et al., 2014; Soofastaei & Fouladgar, 2022a). Despite that, material handling and comminution represent the most energy-intensive aspects, thereby offering significant potential for energy efficiency improvements (Awuah-Offei et al., 2011; U.S. Department of Energy, 2007). Hence, these operations have the greatest potential for boosting energy efficiency and reducing operational costs (Soofastaei & Fouladgar, 2022a).

Several strategies have been identified to improve energy efficiency in mining. These include optimizing mine planning and design, enhancing the efficiency of drilling and blasting (Karpuz, 2018; Lusk & Silva, 2018; Sanchidrián et al., 2018), material handling (Awuah-Offei, 2018a; Awuah-Offei & Frimpong, 2007; Sahoo et al., 2018; Soofastaei et al., 2018), utilizing renewable energy sources (Igogo et al., 2021; J. Sánchez, 2018), and improving processing efficiency (Awuah-Offei, 2018b; Bouchard et al., 2018; Klein et al., 2018; Moats, 2018). Mine planning, for instance, substantially impacts the total energy consumption, as it dictates the volume of overburden removal, the distance materials are transported, and methods used for extraction and beneficiation (Norgate & Haque, 2010).

Advanced Information Technologies (IT), particularly data analytics, are increasingly being adopted to improve mining processes and reduce energy consumption and operational costs (Soofastaei et al., 2017; Soofastaei & Fouladgar, 2022b, 2022a; Trivedi & Fathi, 2021). With mining operations generating significant amounts of data, data analytics and artificial neural networks (ANNs) are beneficial for more effective energy management. Furthermore, ANNs can be employed for predicting and investigating energy efficiency in the mining sector, especially concerning haul truck energy efficiency in surface mining (Soofastaei, 2016). Moreso, researchers have begun deploying Artificial Intelligence (AI) to optimize decision-making and enhance energy efficiency in mining engineering (Soofastaei & Fouladgar, 2022b). Artificial Intelligent models have shown promising results when fed with unstructured and noisy datasets, predicting energy consumption and enhancing energy efficiency.

While technology plays a critical role, the regulatory environment established by the government significantly influences the industry's energy efficiency. Some studies have reported the contradictory effects of these policies on the mining sector (Awuah-Offei, 2016; Henriksson et al., 2014; Hu & Kavan, 2014). Furthermore, carbon taxes and similar regulatory costs can impact the profitability and sustainability of energy-intensive processes (Awuah-Offei, 2016).

In light of the impacts of climate change, the position of renewable energy resources in mining has drawn substantial research interest. For instance, mining locations can be repurposed as renewable energy reservoirs, like low-temperature geothermal resources and on-site renewable energy production can be integrated into the varieties of energy sources (Awuah-Offei, 2016; Carvalho et al., 2014; Hall et al., 2011; Paraszczak & Fytas, 2012; Paredes-Sánchez et al., 2015; Soofastaei & Fouladgar, 2022a; Verhoeven et al., 2014; Watzlaf & Ackman, 2006).

This research is crucial in filling a notable gap in the current mining energy efficiency literature. While there has been considerable investigation into cable shovels and draglines, empirical studies focusing specifically on hydraulic shovels are scarce. Therefore, this research represents a pioneering endeavor to explore the interaction between energy efficiency in hydraulic mining shovels and operator practices, uniquely contributing to the body of knowledge.

This study can potentially reduce energy consumption and lower operational costs in using hydraulic shovels in mining operations by identifying inefficient practices and informing strategies for improved energy use. Moreso, with the robust statistical data analysis approach, this study lays the groundwork for future data-driven decisions that enhance the sustainability and efficiency of mining operations. The outcome of this research could also have important implications for governmental policies and industry training programs, encouraging further investments in data analytics and enabling the development of operator training programs that promote energy-efficient practices.

Furthermore, this study underscores the potential of technology in the mining sector, showcasing the value of statistical data analytics and encouraging further innovation in developing advanced AI systems for sustainable and efficient mining operations. Also, it creates avenues for research that extract actionable decisions from telemetry data beyond what the systems are designed initially to address for hydraulic shovels. In conclusion, this work not only addresses a significant gap in existing literature but also creates opportunities in the mining industry for improved efficiency, policy development, and technological advancement.

2.5. ENERGY EFFICIENCY OF HYDRAULIC SHOVELS

The power and energy of hydraulic excavators are derived from their hydraulic system, thereby transforming the raw mechanical energy produced by the engine into a usable form of hydraulic energy. In most cases, a diesel engine serves as the power generator for the hydraulic mining shovel, which sets the system's mechanisms in motion. Among these mechanisms are hydraulic pumps, which harness the mechanical power generated by the engine and transmute it into hydraulic energy. By generating a full flow, these pumps move hydraulic fluid throughout the system, effectively transferring energy from themselves to the actuators. Once set in motion, the hydraulic fluid can swiftly respond to alterations in flow and pressure, enabling control over the shovel's various parts. The actuators, which include cylinders and motors, are at the receiving end of this hydraulic energy. Through extending and retracting, the cylinders control the movement of the boom, stick, and bucket. At the same time, the motors use the received energy to rotate the superstructure.

As defined earlier in Section 2.4, energy efficiency refers to the proportion of the output of energy to the input of energy. However, in some cases, energy input is defined using proxies (Awuah-Offei, 2018a; Babaei & Hall, 2016). For example, quantifying effective work amid shovel loading operations, encompassing digging and transportation of materials into hoppers or haulage trucks, often proves challenging in a field setting (Abdi-Oskouei & Awuah-Offei, 2014; Awuah-Offei, 2018a). Therefore, the weight of the loaded material or payload is frequently utilized to represent the valuable work done (Equation 2.1). A widely adopted energy efficiency metric for shovel operations among scholars is the energy per unit payload, also known as specific energy. This measure represents the inverse of energy efficiency (Awuah-Offei & Frimpong, 2007; Patnayak & Tannant, 2005). This methodology is rooted in the preliminary research of shovel performance, driven by the aspiration to classify geologic materials or analyze fragmentation outcomes (Acaroglu et al., 2008; Hadjigeorgiou & Poulin, 1998; Iai & Gertsch, 2013; Muro et al., 2002; Scoble & Muftuoglu, 1984).

$$\eta = \frac{P}{E_t}\left(\frac{tonnes}{kWh}\right) \tag{2.1}$$

Specific energy was initially conceived to determine the excavation difficulty of material or muck piles, not to assess the efficiency of the shovel's performance (Awuah-Offei, 2018a). However, it has become commonplace for academics and professionals to employ payload per unit energy consumed[5] or specific energy as a yardstick for loaders' (shovels) energy efficiency. This routine, however, presumes that the shovel's loading speed does not factor into defining efficient loading.

[5] A definition closer to the theoretical concept of energy efficiency.

Contrarily, mining leaders and professionals attach significant importance to shovels' rate of loading. as it dictates the overall rate of production of the material handling system. Therefore, it can be inferred that assessing the shovel operations' energy efficiency is more appropriately done by employing the energy expended for each unit rate of loading (Awuah-Offei, 2018a; Awuah-Offei & Frimpong, 2007). Babaei Khorzoughi and Hall (2016) demonstrated that a shovel could use drastically different energy amounts while digging at comparable loading rates. Some other authors have suggested that loading's energy efficiency should be perceived as the proportion of the loading rate to the input of energy (Awuah-Offei, 2018a). Therefore, this thesis uses this definition of energy efficiency of loading (Equation 2.2).

$$\eta_l = \frac{E_t}{P}\left(\frac{kWh}{tonnes}\right) \tag{2.2}$$

2.6. FACTORS AFFECTING ENERGY EFFICIENCY OF HYDRAULIC SHOVEL

The efficiency of a hydraulic shovel is intrinsically linked to two core parameters: payload or productivity and energy consumption, as Equation (2.1) implies. Therefore, we must explore and comprehend the various elements influencing these key parameters to effectively manage and potentially enhance its energy efficiency. A review of prior research on dragline and cable shovels reveals that four primary factors govern these parameters (Figure 2.5): operator practices, mine design and planning, conditions of operation, and attributes of equipment (Abdi-Oskouei, 2013; Awuah-Offei, 2016, 2018a;

Awuah-Offei et al., 2011; Hettinger & Lumley, 1999; Kecojevic et al., 2011). Past research and studies (even though mostly on cable shovels) have been instrumental in shedding light on these influencing factors, thus forming the basis of our understanding and providing guidance for further investigation.

Mining shovel activities play a central role in energy efficiency within mining practices, with a particular focus directed towards the excavation phase (Awuah-Offei, 2018a). This process is highly energy-intensive; thus, models of its kinematics and dynamics have been the primary research focus (Awuah-Offei, 2018a; Guzman et al., 2015). The energy consumed during excavation primarily relies on the crowd and hoist forces, the speeds associated with them, and the duration of the digging process (Awuah-Offei, 2016; Awuah-Offei & Frimpong, 2007, 2011). In addition, these forces must be strategically deployed to navigate through the prescribed digging trajectory and overcome resistance presented by the material to be excavated (Awuah-Offei & Frimpong, 2007).

The concept of shovel energy efficiency revolves around two primary parameters: the energy consumed during loading and the rate of loading (Awuah-Offei, 2018a; Awuah-Offei & Frimpong, 2007). Consequently, any factor influencing these parameters will directly impact energy efficiency (Awuah-Offei, 2018a). Previous studies have pinpointed conditions of operation, shovel attributes, mine design, operator skills, and practice (Figure 2.5) as crucial elements impacting energy efficiency (Abdi-Oskouei, 2013; Awuah-Offei, 2016; Awuah-Offei et al., 2011).

Energy efficiency is the proportion of per unit time productivity in relation to the input of energy (Equation 2.1), which varies based on whether the loading equipment is electrically or diesel-powered. For instance, for non-electric (diesel) excavators like hydraulic shovels, the fuel volume serves as a proxy for energy input. As such, the excavation specifications, mine planning and design, and conditions of operation will dictate the production rate. At the same time, the energy input will be influenced by factors such as the load of the material, the bucket trajectory[6], and digging resistance.

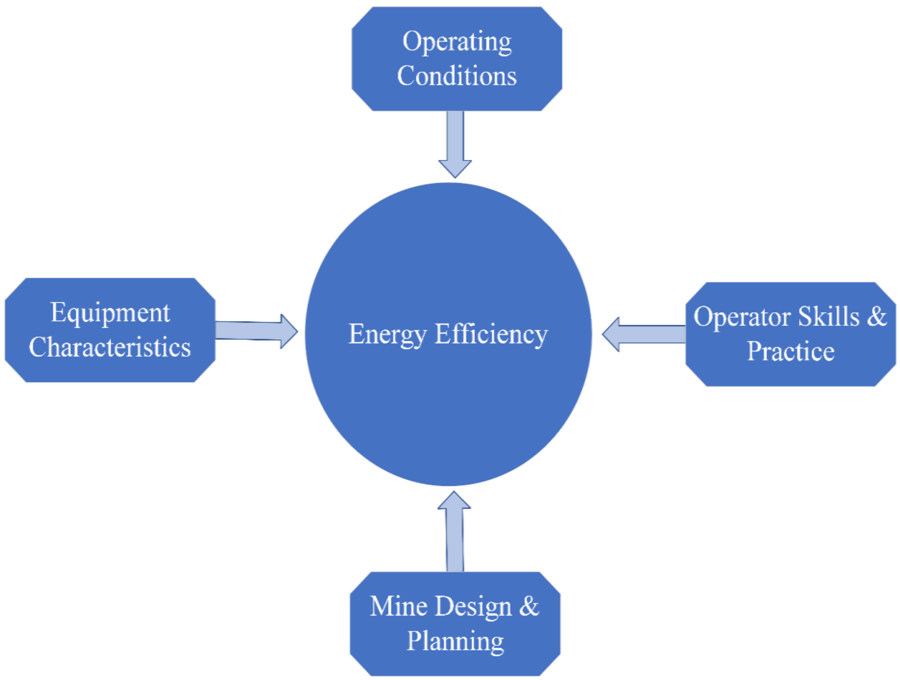

Figure 2.5 Factors affecting shovel energy efficiency

Whether naturally occurring or design-induced, operating conditions can significantly impact the energy efficiency of mining shovels. Factors such as resistance to

---

[6] Bucket's trajectory refers to the path or route that the bucket of an excavator, shovel, or dragline follows during the operation of loading or digging. It includes the movement of the bucket from when it starts digging into the material, lifts the material, swings towards the dump or hauler, releases the material, and finally swings back to the starting point for the next cycle.

digging, bench profiles, and truck matching all play crucial roles in this context (Abdi-Oskouei, 2013; Awuah-Offei et al., 2011; Awuah-Offei & Frimpong, 2007; Karpuz et al., 1992). Sub-optimal ground fragmentation can lead to challenging digging conditions and, as a result, causes an increase in crowd and hoist forces, leading to higher energy consumption and reduced energy efficiency (Awuah-Offei, 2018a; Singh & Narendrula, 2006). Similarly, inefficient truck matching can result in wasted energy as the shovel waits for the truck's arrival, consequently diminishing energy efficiency (Awuah-Offei, 2018a).

Due to the limited research on hydraulic shovels, most of the literature reviewed in this thesis is focused on cable shovels. Considering the operational similarities between these shovels, this author hypothesizes that the findings from cable shovels can be extrapolated and applied to hydraulic shovels. However, it is important to note that no empirical study is available to back this up. As we explore these factors, it is also crucial to consider the potential transference of knowledge between these two types of heavy-duty machinery. Thus, previous research on cable shovels may be more relevant to our understanding of hydraulic shovel energy efficiency than initially perceived.

2.7. INFLUENCE OF THE PRACTICES OF OPERATORS ON SHOVEL'S ENERGY EFFICIENCY

The influence of the practices of operators on mining shovels' energy efficiency is critical and cannot be overstated (Awuah-Offei, 2016, 2018a; Awuah-Offei et al., 2011). The role of an operator extends beyond the operation of the machines. Energy efficiency is significantly influenced by the skills and practices of operators, even when operating conditions are optimal (Abdi-Oskouei & Awuah-Offei, 2016; Awuah-Offei & Frimpong, 2007; Lumley & Lumley, 2005; Sahoo et al., 2018). Various elements of operation, such as cycle duration and fill factor of the bucket, are controlled by the operator, thereby directly influencing the energy consumption and rate of production (Awuah-Offei, 2016). It is important to note that the trajectory significantly affects the fill factor and energy consumption (Awuah-Offei, 2016).

Operators' role, particularly their skill level, and practices, is crucial to energy efficiency (Awuah-Offei, 2016). The operator influences the payload and cycle duration, which are directly tied to energy efficiency (Awuah-Offei & Frimpong, 2007, 2011). The travel rate and operator-selected trajectory influence the duration of the cycle. At the same time, the payload depends on the operator's effectiveness in filling the bucket. Higher travel rates could potentially reduce the cycle time, increase the loading rate, and decrease overall energy consumption per cycle (Awuah-Offei, 2016).

The operator's role in cable shovel operations has also been highlighted in the relationship between hoist speed and faster digging (Awuah-Offei, 2016; Patnayak et al., 2008). Increasing the crowd speed can result in deeper cuts and, consequently, higher energy consumption. However, the loading rate can be optimized to decrease overall energy consumption (Awuah-Offei & Frimpong, 2007). This balance, achieved through effective operator skills and practices, can substantially boost the energy efficiency of mining shovel operations.

Research has indicated significant inefficiencies in energy usage due to the practices of operators (Abdi-Oskouei & Awuah-Offei, 2014, 2016; Kecojevic et al., 2014; Patnayak et al., 2008). Some operators, when compared to the most efficient operators, could expend up to 40% more energy for each tonne of materials produced

(Awuah-Offei, 2016). Furthermore, it has been shown that the execution speed and trajectory used by the operator during the excavation phase were the most impactful elements influencing energy consumption (Awuah-Offei & Frimpong, 2007; Wei & Gao, 2012).

Interestingly, greater depths of cut could increase the fill factor, which potentially improves productivity and energy efficiency. However, deeper cutting depths are generally associated with increased energy consumption (Awuah-Offei & Frimpong, 2007; Karpuz et al., 1992; Patnayak et al., 2008). Consequently, striking a balance between the depth of cut and the fill factor can significantly influence energy efficiency (Awuah-Offei, 2016). In addition, the pace at which an operator carries out a trajectory is in direct ratio to the power demand, resulting in decreased cycle duration and increased productivity (Awuah-Offei, 2016). Research has shown this to be the case for cable shovels, and this author believes this will also be generally true for hydraulic shovels because of the similarities in the mode of operation of both mining shovels. While the author believes this is true for hydraulic shovels, the relationship between speed, power draw, and productivity is not always linear. An optimum speed may result in the highest productivity without causing excessive wear on the machine or inefficient use of power. It is essential to note that this relationship will also be influenced by various other factors, such as the properties of the material being moved, the efficiency of the hydraulic system, the operator's skill and experience, and the machine's overall condition.

However, simulation experiments suggest that an attempt to increase the digging speed requires careful optimization to maintain overall energy efficiency (Awuah-Offei & Frimpong, 2007). In addition, the interaction between the operator and other equipment units may introduce inefficiencies, resulting in delays and periods of inactivity. Although this aspect has not been extensively investigated, it is clear that a less-than-optimal interplay between the operator and equipment can cause a drop in energy efficiency. (Awuah-Offei et al., 2011).

In conclusion, the practices of operators substantially influence shovels' energy efficiency in mining. While the cable shovel research provides some basis for us to hypothesize that operators will have a similar impact on hydraulic shovel energy efficiency, there is no empirical evidence to support this. This area of research should be explored to improve hydraulic shovel energy efficiency. More nuanced research exploring the optimization of these practices, including operator interactions and execution speeds, could yield substantial energy savings and productivity gains. This underscores the need for systematic and rigorous training of operators to ensure that their practices align with energy-efficient and sustainable mining operations.

2.8. SUMMARY

Integrating telemetry and sensing technologies in the mining industry, particularly excavators, has significantly improved operational efficiency, safety, and cost-effectiveness. These advanced systems enable real-time equipment monitoring, enhancing its utilization and maintenance while promoting worker safety. Predictive maintenance can foresee potential equipment failures, minimizing downtime and expenses. Thus, telemetry supports more informed decision-making, fostering overall performance in mining operations.

We can gauge a shovel's energy efficiency through factors like payload and overall energy expenditure, indicative of the accomplished work and input energy. Boosting energy efficiency provides a viable method for catering to growing energy needs while lessening the environmental footprint of energy use. The hydraulic shovel's productivity, flexibility, and high performance have made it a prevalent piece of equipment in surface mines. Monitoring systems installed on these shovels can furnish critical insights, enable data-driven decision-making, and create strategies for improving equipment performance in mining operations. Despite these benefits and strategies, past work focused on dragline and cable shovel energy efficiency. There has not been any empirical study that has explored the influence of the practices of operators on hydraulic shovel energy efficiency.

Identifying the factors influencing the productivity and energy usage of a hydraulic shovel is vital for effectively managing its energy efficiency. Performance indicators integral to shovels' energy consumption and productivity include the factor of fill, duration of the cycle, the energy of digging, and the positions where the shovel engages and disengages. The practices of operators, operating scenarios, mining design and planning, and the attributes of the equipment govern these indicators. Modifying operator performance is the most cost-effective way to boost energy efficiency among these factors.

While altering the operating conditions may not always be feasible in a mine, optimizing the shovel drive mechanism can be expensive, and assigning inefficient tasks to shovels can sometimes be inevitable. Investing in operator training to boost their performance is a comparatively affordable enhancement and an excellent approach to improving energy efficiency. Therefore, comprehending the influence of operators' practices on hydraulic shovel energy efficiency and quantifying this connection is crucial. Thus, recognizing variables that differ between operators can enhance training systems.

3. HYDRAULIC SHOVEL PERFORMANCE DATA COLLECTION AND DEVELOPMENT OF CYCLE SAMPLING ALGORITHM

This section of the thesis describes the data acquisition, data structure, pre-processing, development of cycle and sub-cycle sampling algorithms, and algorithms to extract cycle-based information. All the data used in this research came from a monitoring system installed on a 40.5 yd^3 bucket hydraulic shovel. The work uses this data to develop algorithms to sample cycles (and identify sub-cycles) and extract cycle-based information for the key performance indicators needed for this study.

3.1. DATA ACQUISITION AND PRE-PROCESSING

Typically, monitoring systems on excavation machinery gather and archive a wide range of real-time operational parameters, including vital performance indicators such as payload, cycle time components, voltage, angles, pressure, and temperatures. To effectively measure energy efficiency, it is necessary to track the elements of energy usage throughout the hydraulic shovel cycle. Such elements include cylinder pressures, torque, and angular displacements that can be used to estimate energy input.

The data used in this work is acquired from a 40.5 yd^3 bucket Hitachi EX-5600 hydraulic shovel operated by 16 different operators in a one-month operation period using a telemetry-based commercial monitoring system. The on-site experiment consisted of a month-long observation of the same shovel operated by different operators under similar conditions and data downloaded via telemetry for the purpose of this research. The commercial machine monitoring systems with sensors installed on the hydraulic shovel collected real-time information, including roll, pitch, yaw angles, accelerations, and velocities in x, y, and z directions, for the cab, bucket, stick, and boom. The system also collects boom head and rod pressures. This researcher worked with the commercial monitoring systems owner to download the data in comma-separated values (csv) format. This raw data, contained in the sensor download files, require further processing into meaningful information that can be used for this study. The systems owner provided a MATLAB script that this researcher modified to read the information and process the data into the initial data set for this research (Figure 3.1).

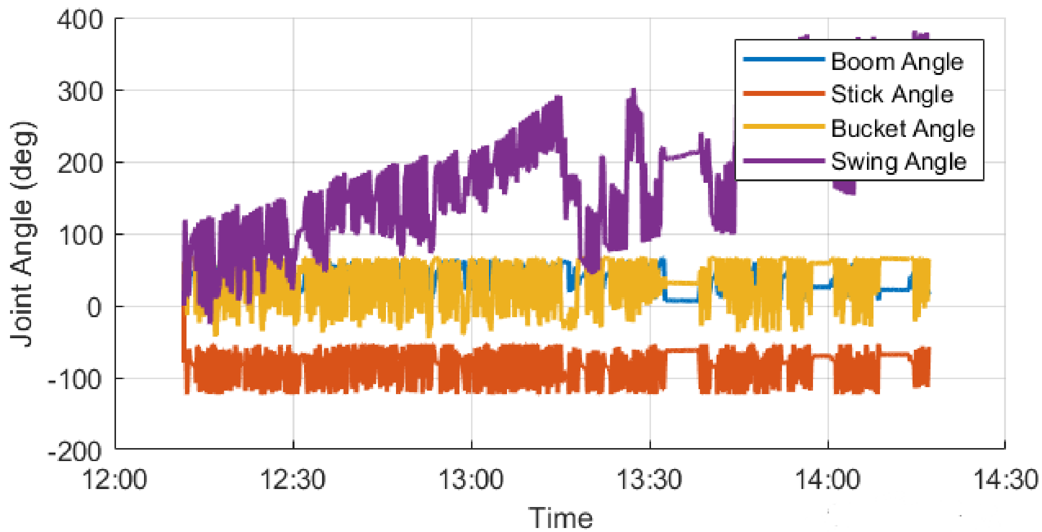

Figure 3.1 Sample plot of angular displacements from the initial data from a single sensor file

After the sensor data is pre-processed (Figure 3.1), the output data includes twenty-six variables sampled at 30 Hz frequency. For the purpose of this study, ten of these variables are used to elucidate the relationship between operators' practices and the energy efficiency of hydraulic shovels. The selected variables include timestamps, payload (which is estimated from the sensor data using the monitoring system's proprietary algorithm), boom, swing, bucket, and stick angular displacements, bucket height (estimated from the sensor data using the shovel kinematics), state enum (a variable representing the state of the shovel during the cycle generated using the commercial partner's proprietary algorithm), operator id, and torque. This research uses the monitoring data of these 10 variables sampled at 30 Hz to estimate the key indicators for each cycle used for the analysis in this work.

3.2. CYCLE SAMPLING ALGORITHM AND VALIDATION

An essential step in processing performance data involves isolating 'loading cycles' or 'duty cycle' segments related to loading so that this research can examine payload per cycle and energy consumed per cycle to evaluate energy efficiency. Thus, establishing the loading cycle is crucial for advanced analysis. However, reliably identifying these loading cycles is not trivial. In its most simplified form, a basic hydraulic shovel duty cycle consists of swinging into position, digging to fill the bucket with material, swinging out, and finally dumping the materials to begin the next cycle (Figure 2.4).

However, the duty cycle tasks are more intricate in reality as two or more of the previously mentioned activities often coincide. For instance, the dipper might start swinging toward the truck while it is still in the digging phase. Moreover, additional shovel activities such as cleaning the face or loosening material complicate the movement of the dipper. Lastly, there are two directions the shovel operator can swing. However, most of the cycles are cycles where the shovel operator swings in the direction where his vision of the truck is unobstructed by the boom. Thus, the challenge lies in processing and interpreting the shovel performance data to identify those shovel cycles that are useful for comparing operator energy efficiency (i.e., it is not desirable to compare a cycle that includes bench cleanup with another that does not). Thus, the goal of this research was to sample the "ideal" cycles (i.e., cycles with clear demarcation of stages, no bench clean up, and where the swing is to the direction where the operator's vision is not unobstructed by the boom) for each operator so that our comparison is a direct comparison.

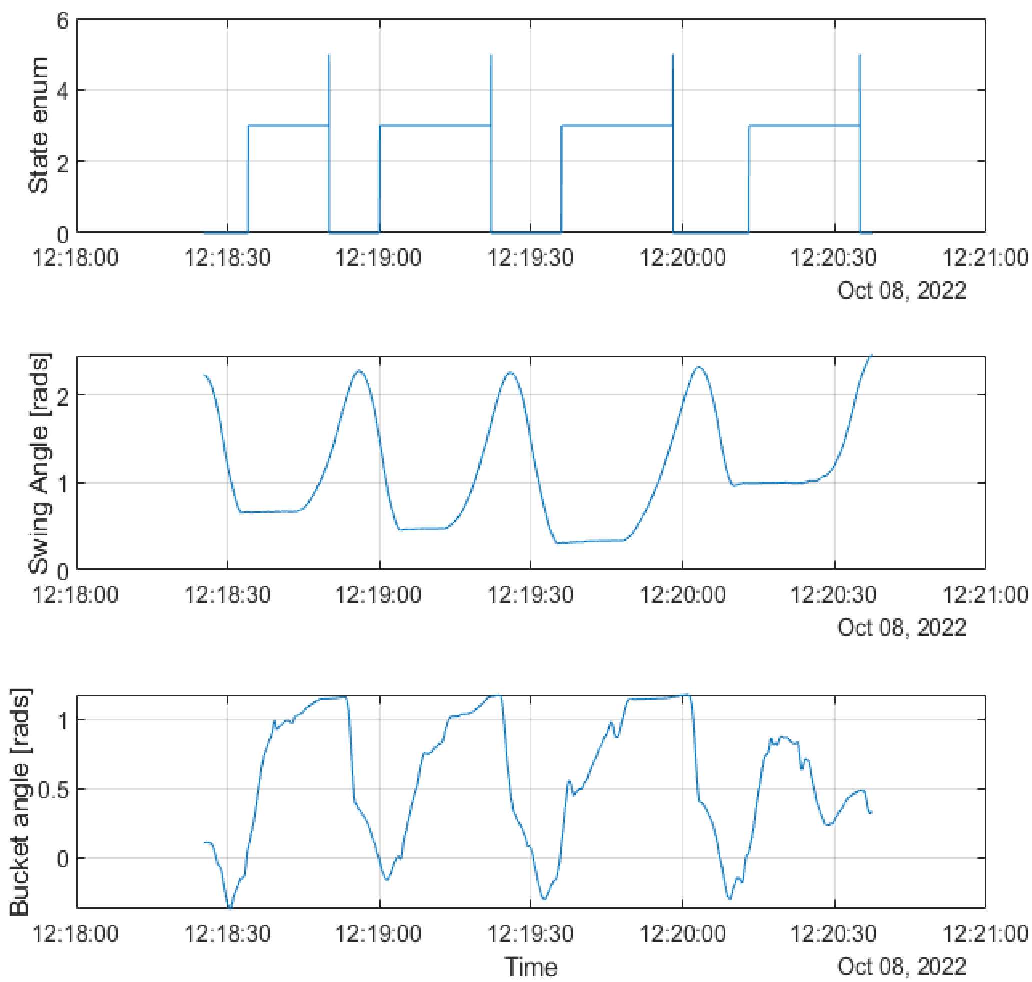

Figure 3.2 Sample plots of state enum, swing angle, and bucket angle

The pre-processed data from using the MATLAB script from the commercial partner (modified by this researcher) contains the state enum variable, which is a variable that denotes the state of the hydraulic shovel (e.g., state enum = 1 means swinging and state enum = 2 means digging, state enum = 5 means dumping). However, the commercial partner recognizes deficiencies in the performance of their algorithm. For example, as shown in Figure 3.2, state enum values of 5 (dumping) do not coincide with the end of the cycle (when swing angle is at the peak) as it should.

Consequently, this research needed to develop its own algorithm to identify cycles for the sampled cycles. However, the availability of the state enum variable was crucial to the developed algorithm. The state enum data allowed this researcher to sample loading cycles effectively, facilitating a clearer understanding of the shovel's operation and performance.

One of the objectives of this study is to develop algorithms to sample portions of the hydraulic shovel monitoring data corresponding to individual loading cycles that facilitate comparison of operator performance. Once the loading cycles have been accurately sampled, it becomes possible to estimate the useful information per cycle that can be used to estimate energy per unit loading rates and variables hypothesized to explain operator performance differences.

The algorithm developed is a comprehensive tool for sampling loading cycles and their corresponding sub-cycles in a given data set from the particular monitoring system used in this research. This researcher believes, however, that the algorithms can be extended to similar telemetry data for hydraulic shovels that monitor the swing and bucket angular displacements. The algorithm (implemented in MATLAB as a function) requires three arguments as input: the state enum variable and swing and bucket angular displacements. Figure 3.3 shows the logic of the algorithm.

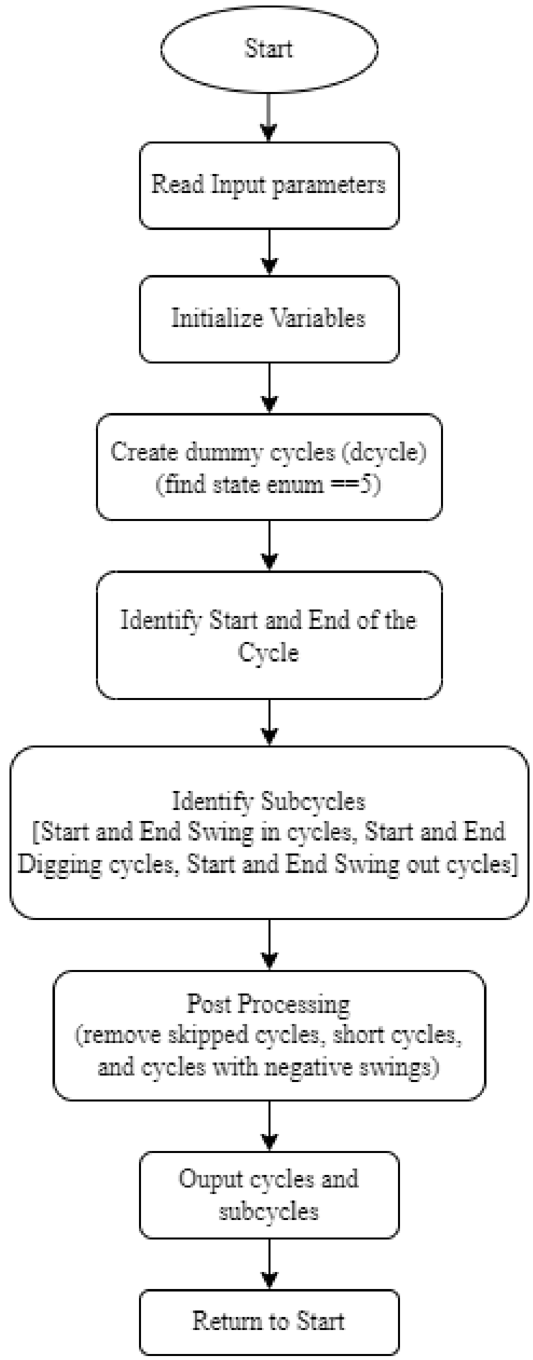

Figure 3.3 A simplified flowchart illustrating the cycle sampling algorithm

The function initializes variables and partitions the data into cycles by creating "dummy" cycles (dcycles). This partitioning is achieved by identifying where the state enum equals 5 (corresponding to the instant when the bucket dumps according to the proprietary algorithm), using these points to delineate sections corresponding to each dummy cycle. However, as shown in Figure 3.4, these cycles do not align with the actual end of the cycle. For example, a careful examination of the plots in Figure 3.4 shows the reader that the machine is still swinging significantly at the beginning and end of these dummy cycles.

However, using the partitioned data in the dummy cycles, this work developed an algorithm to detect a more realistic start and end of the cycle. The function detects the actual start of the cycle as the peak of the swing angle signal in a section of the signals where the state enum equals 2 for dummy cycle i and dummy cycle $(i+1)$. The peak of the swing angle represents the stationary point in the signal where the operator reverses swinging at the end of the cycle.

In essence, the signal between the state enum 2 of the first dummy cycle and the following dummy cycle's state enum 2 is extracted, the peak point within this interval is identified, and this point is stored as the start of the cycle. It is important to note that the first cycle's start is always identified by locating the forward peak from the state enum = 2 of the first dummy cycle to the state enum 2 of the next dummy cycle. This is based on the assumption that the first peak of the Swing Angle (Figure 3.4) represents only part of a cycle, with the actual complete cycle beginning from the subsequent peak.

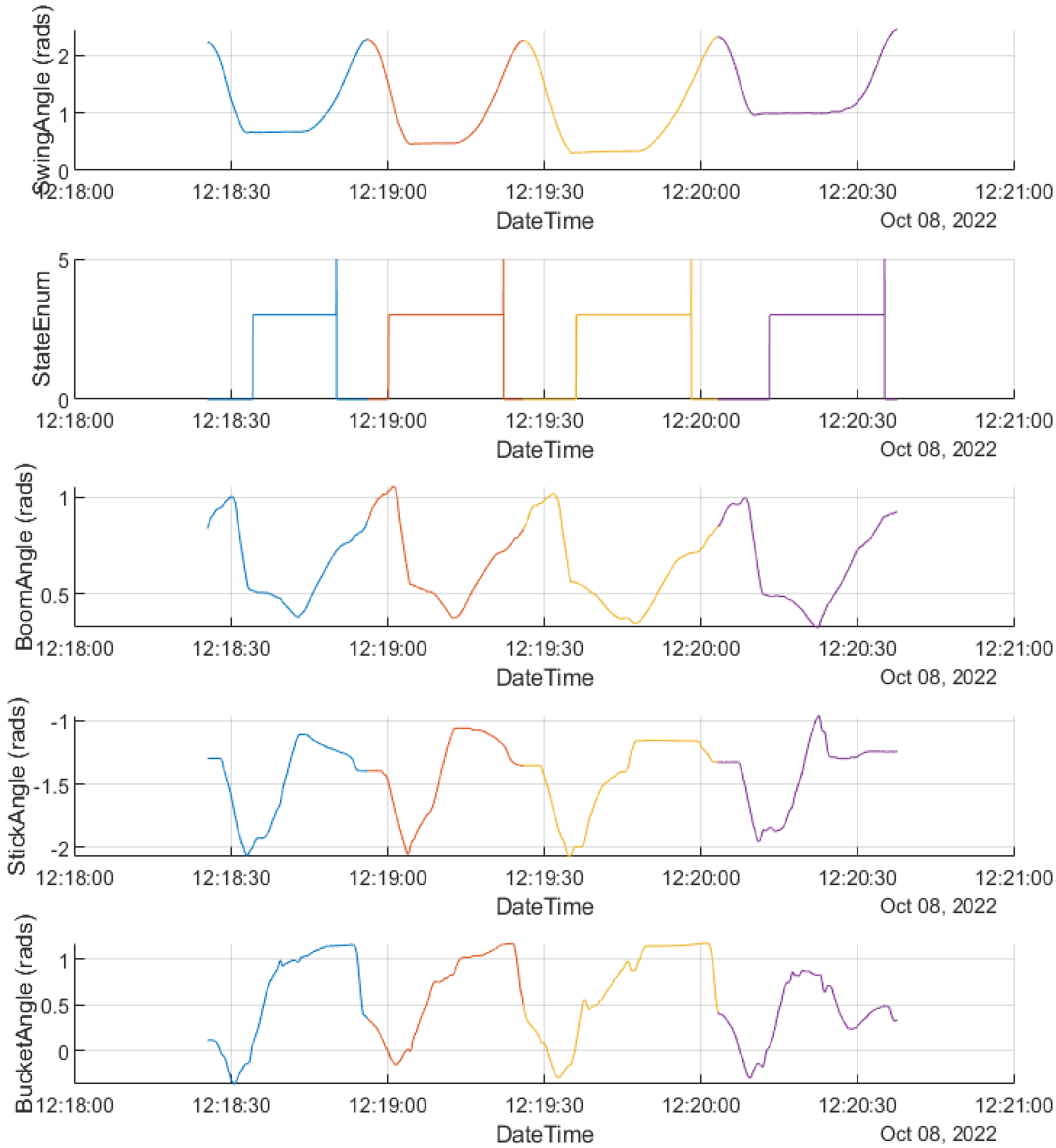

Figure 3.4 Sample plots of generated dummy cycles

Furthermore, to identify the end of the cycle, the function moves forward to find the peak point of the Swing Angle signal between the state enum = 5 of dummy cycle i and the state enum = 2 of dummy cycle $(i+1)$ (Figure 3.4). This is under the premise that the next peak would occur after state enum = 5 of the dummy cycle $(i+1)$. These peaks define the start and end of each cycle. In summary, the function generates dummy cycles based on the state enum variable provided by the commercial partner. Using the data derived from these dummy cycles, it segments the signals into distinct portions to facilitate the identification of the start and end of each cycle, which are determined by the first and second peaks, respectively.

Upon pinpointing the start and end of the cycles, the function next determines subcycles. This work determines three subcycles: swing-in, digging, and swing-out cycles. So, the first thing the function does in this phase is to determine the conclusion of the swing-in phase, as the swing-in cycle starts at the beginning of the cycle. It does so by finding the point of minimum value of the bucket angle signal within the interval between the start and end of the cycle. The researcher has concluded from observing multiple signals and evaluating the kinematics of the hydraulic shovel that the bucket angle reaches a minimum at the initiation of digging (Figure 3.4), signifying the end of the swing-in phase and the commencement of the digging phase. In essence, the point of minimum bucket angle marks the transition between the swing-in and digging cycles.

To identify the end of the digging cycle, the researcher employs a heuristic approach to assess the flatness of the digging cycle (Figure 3.4). Ideally, the swing angle remains constant during digging (Figure 3.4) and begins to increase when digging ends. The challenge, however, lies in pinpointing the moment when the swing angle starts to rise (Figure 3.4), which denotes the end of the digging cycle. To this end, the function employs heuristics to find points within the interval spanning from the start of the digging cycle to 75% of the duration from the start of the digging cycle to the end of the cycle, where the difference in swing angle for a time step (1/30 seconds) is less than 10^{-4} rads.

The choice of 75% of the duration is made to optimize computational efficiency and is based on the assumption that this is more than enough to capture the beginning of the swing-out phase. Once a point is found, the function finds all the points that have a swing angle that is less than 10^{-4} of the previous point and identifies the points in the range where the swing angle jumps higher than 10^{-4} after a series of points where the swing angle is nearly flat. The function then tests to see if the absolute value of the swing angle difference between that point and the start of the digging cycle is more than 0.15 rads. The end of the digging cycle is that point with a higher difference in swing angle of greater than 0.15 rads that follows a series of points with incremental swing angles of less than 10^{-4}. The author determined these threshold values through a series of experiments and conducted a sensitivity analysis to ensure these are the “optimal” threshold values.

Inherently, the conclusion of the digging cycle initiates the swing-out cycle, whose end is congruent with the end of the cycle. Lastly, the function carries out a series of validations, eliminating skipped cycles, cycles with exceptionally short subcycles (i.e., those below 1.67 secs), and cycles where the operator swings in the negative direction. Figure 3.5 shows the sample results of the algorithm.

Additionally, this author used visual inspections of the swing, boom, and stick angular displacements plots and the state enum variable (Figure 3.5) to validate the cycle sampling algorithm. The objective was to visually inspect these plots to confirm if they conformed to the ideal loading cycle pattern. The validation exercise used a set of sensor data, including 15 sensor files, for verification. Figure 3.5 shows a visual illustration of the validation for the first of the 15 files, and all the results of the validation exercise are further presented in Table 3.1.

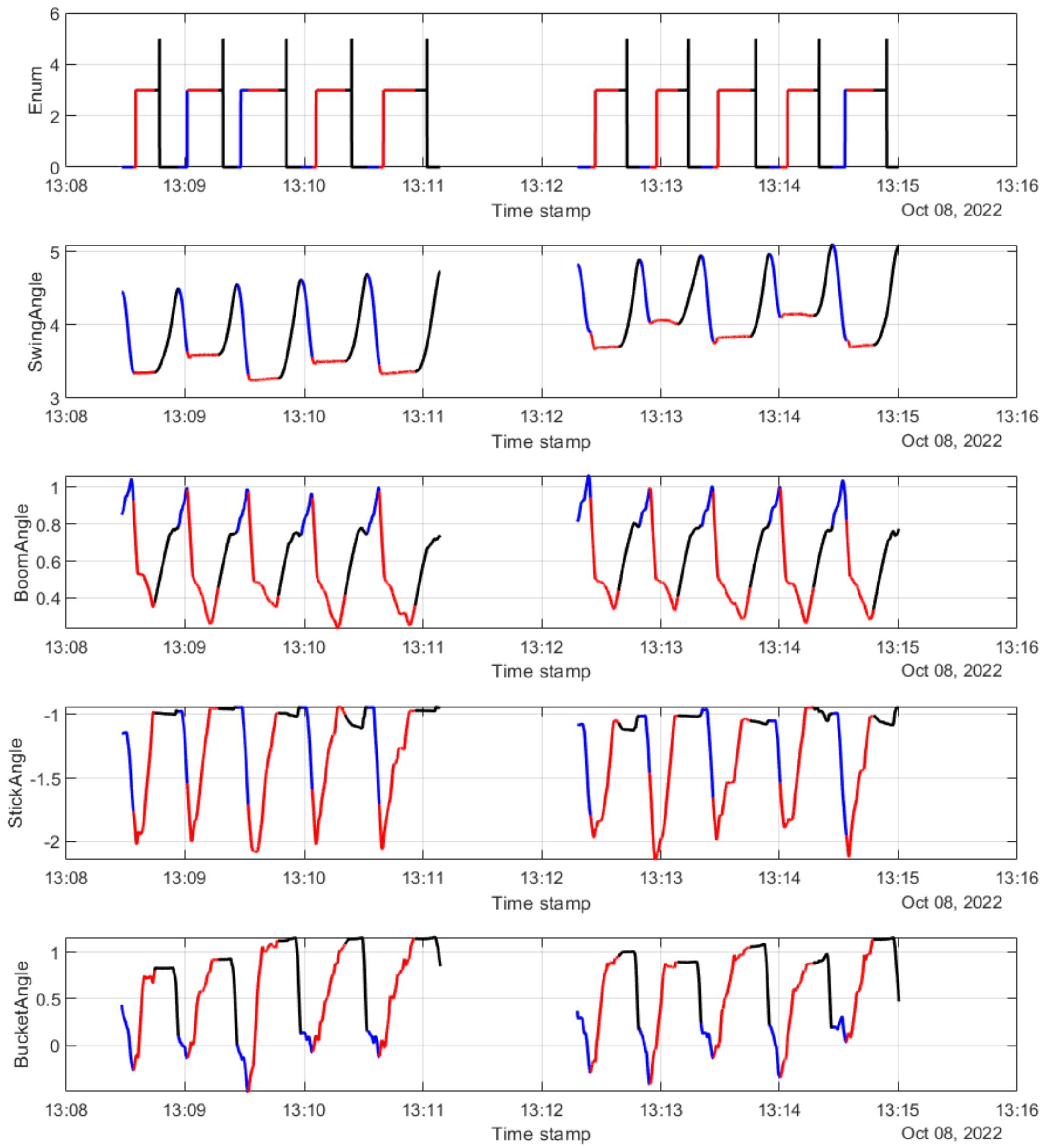

Figure 3.5 Plot of the output of the cycle sampling algorithm for validation (blue = swing in; red = digging; and black = swing out subcycles)

Table 3.1 shows that the cycle sampling algorithm is 98% accurate in identifying the type of cycles the algorithm was designed to sample. This work deems this level of accuracy acceptable. Notably, the primary aim of developing this algorithm was not to identify every possible loading cycle. Instead, the focus was pinpointing those cycles corresponding to “ideal” loading.

Table 3.1 Cycle sampling output validation results

Sensor File #	No of Cycles Sampled	No of valid cycles	Accuracy
1	68	66	97%
2	33	31	94%
3	17	17	100%
4	0	0	100%
5	4	4	100%
6	0	0	100%
7	17	15	88%
8	81	80	99%
9	22	22	100%
10	35	33	94%
11	72	71	99%
12	75	75	100%
13	62	62	100%
14	0	0	100%
15	48	47	98%
Total	534	523	98%

In conclusion, the cycle sampling algorithm is a robust tool that accurately samples loading cycles and identifies their sub-cycles from the sensor data. It is useful for further work to extract cycle-based information that can be used to estimate energy per unit loading rate and variables that could help explain operator differences.

3.3. ALGORITHMS TO EXTRACT CYCLE-BASED INFORMATION

The principal objective of this research is to develop robust algorithms capable of extracting valuable insights from raw shovel performance data and transforming it into data per cycle. These insights are crucial for understanding the impact of operator behaviors on energy efficiency and productivity. After successfully devising the cycle sampling algorithm, the next pivotal step is to use it as a basis for further algorithm development. This stage of the research involves the creation of numerous algorithms, each designed to tap into a specific key performance indicator (KPI). Based on the nature of the monitoring data and a review of the literature focused on finding factors that are likely to explain differences in operator energy per unit loading rate, this researcher selected eight KPIs for each cycle: cycle time and cycle time components, payload, energy use, dump height, and boom, swing, and stick angles.

Additionally, this work developed an algorithm to extract operator identities to facilitate separating data by operator. All these algorithms are implemented in MATLAB as MATLAB functions. Finally, the work developed a high-level algorithm in a MATLAB function that reads the input files and calls all the necessary functions to generate the output. The following sections provide a detailed account of these subsequent algorithms, illustrating how they extract relevant information from the raw sensor files and ultimately contribute to this study. All the algorithms use the output of cycle sampling and pre-processing algorithms.

3.3.1. Cycle Time Components Algorithm. The cycle time components algorithm computes the time duration for each cycle and its subcycles: swing-in, digging, and swing-out. The inputs of the function are **c**, which is an $n \times 2$ matrix (where n is the number of cycles in the data file) containing the row numbers in the sensor data that mark the start and end of a cycle (Figure 3.6), **s** is an $n \times 4$ matrix containing the row numbers in the sensor data that mark the start and end of the swing-in, digging, and swing-out sub-cycles (Figure 3.6), and **t** is a vector containing the timestamps from the sensor data. The outputs are vectors containing, in seconds, containing the duration of the cycles and subcycles.

$$\mathbf{c} = \begin{bmatrix} 7{,}641 & 8{,}720 \\ 12{,}254 & 13{,}187 \\ 13{,}187 & 14{,}090 \\ 14{,}090 & 15{,}221 \\ \mathrm{M} & \mathrm{M} \\ 209{,}967 & 211{,}047 \end{bmatrix} \qquad \mathbf{s} = \begin{bmatrix} 7{,}641 & 7{,}886 & 8{,}174 & 8{,}720 \\ 12{,}254 & 12{,}414 & 12{,}797 & 13{,}187 \\ 13{,}187 & 13{,}353 & 13{,}692 & 14{,}090 \\ 14{,}090 & 14{,}298 & 14{,}759 & 15{,}221 \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} \\ 209{,}967 & 210{,}145 & 210{,}675 & 211{,}047 \end{bmatrix}$$

Figure 3.6 Sample input data for cycle time components algorithm

Initially, the function identifies the start and end timestamps of each cycle by indexing the time array **t** with the start and end indices contained in columns 1 and 2 of **c**. The difference between these timestamps gives the total cycle time in seconds. The function follows a similar process for each subcycle: swing-in, digging, and swing-out. The start and end timestamps of each subcycle are identified by indexing the time array **t** with the relevant indices from **s**. The duration of the swing-in, digging, and swing-out cycles are calculated as the difference between the end and start timestamps, again in seconds. By breaking down the total cycle time into these subcycle durations, the function provides more detailed insight into the operational efficiency and potential areas for optimization in the loading process.

3.3.2. Payload Algorithm. The payload algorithm (developed as a MATLAB function) computes the final payload for each cycle. It takes as input two arguments: **P** and **c**. **P** represents the payload data sourced from the data file, and **c** is an $n \times 2$ matrix where n corresponds to the number of cycles in the data file (Figure 3.6). This matrix c contains the row numbers from the sensor data that mark the start (column 1) and end (column 2) of each cycle (Figure 3.6). The output **FP** is a vector containing the final payload for each cycle. The function calculates **FP** within each cycle by finding the maximum payload value between the start and end indices as defined by the corresponding row in **c**. This maximum value is then assigned to the corresponding index in **FP**.

The function also checks if the computed payload value exceeds the maximum allowed value of 120 tonnes. Equation 3.1 succinctly shows what the algorithm does, where $\mathbf{P}_{\mathbf{data}}$ is the vector of payloads for that cycle. This capping is necessary because the bucket/shovel used in this context cannot physically load more than 120 tonnes of material. However, due to anomalies in the proprietary algorithm of this researcher’s commercial partner, some payload estimates might incorrectly exceed this limit. In such cases, the payload value for that cycle is set to the limit of 120 tonnes.

The final payload, **FP**, is thus a vector representing the maximum payload for each cycle, with all values capped at 120 tonnes. This computation provides an understanding of the payload handling capacity per cycle, allowing for a more accurate assessment and enhancement of the operational efficiency of the loading process.

$$\text{Payload} = \min\left(120, \max\left(\mathbf{p}_{\text{data}}\right)\right) \tag{3.1}$$

3.3.3. Energy Algorithm. The energy algorithm calculates the work done by the boom (also termed the boom's energy) for each cycle. It takes three inputs: **Γ**, **BoomAngle**, and **c**. **Γ** represents the torque data retrieved from the data file. **BoomAngle** refers to the angle data of the boom for each data point. **c** is an $n \times 2$ matrix, where n denotes the number of cycles in the data file. The output **E** is a vector containing the boom's work, measured in kilojoules (kJ), for each cycle.

The basic premise of this function is that energy (workdone) is the product of torque and angular displacement. The function calculates **E** such that within each cycle, it first computes the absolute difference in the **BoomAngle** from the start to the end of the cycle as defined by the corresponding row in **c**. This value, denoted as **θ**, represents a vector containing the change in the boom angle during the cycle (i.e., the boom's angular displacement). Next, the function computes the scalar product (dot product) of **θ** and the corresponding **Γ** values (torque data) between the start and end indices of the cycle (Equation 3.2). This product represents the work done by the boom during the cycle, which is then assigned to the corresponding index in E.

$$\text{Energy} = \sum_{i=1}^{n} \left(\Gamma_i \times \theta_i \right) \tag{3.2}$$

Finally, the function divides all elements in **E** by 1,000 to convert the energy measurements from Joules to Kilojoules (kJ). By computing the energy spent in each cycle, this function provides a crucial measure of the boom's operational efficiency and energy consumption during the loading process.

3.3.4. Boom Angle Algorithm. The boom angle function calculates the range of boom angles associated with each cycle in the data file. It requires two inputs: **c** and **BoomAngle**. **c** is an $n \times 2$ matrix, where n corresponds to the number of cycles in the data file. **BoomAngle** is the vector containing the boom angle data from the data file. The output, **BA**, is an $n \times 1$ vector that contains the range of boom angles for each cycle.

In each cycle, defined by a row in **c**, the function identifies the segment of the **BoomAngle** data corresponding to that cycle. It then calculates the minimum and maximum values of the **BoomAngle** within this segment (Figure 3.7). The difference between the maximum and minimum angles gives the range of boom angles for that cycle, which is stored in the corresponding index in **BA**.

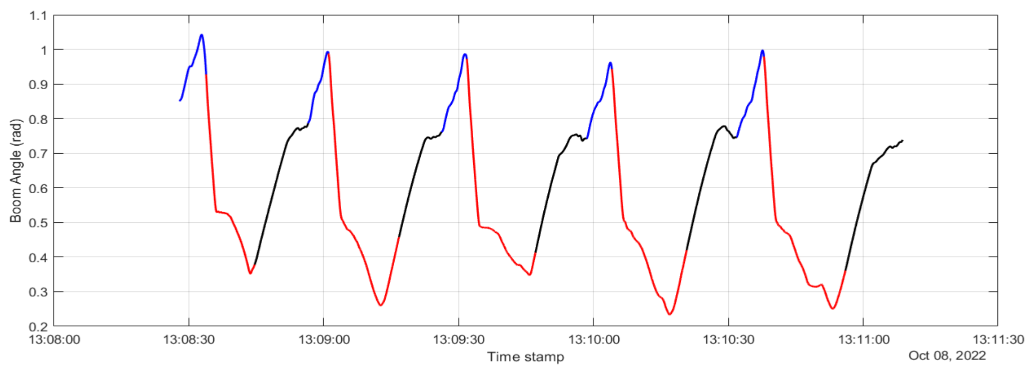

Figure 3.7 Structure of “typical” boom angles during a cycle (blue = swing in; red = digging; and black = swing out subcycles)

By generating the range of boom angles for each cycle, this function offers valuable insights into the scope of boom movement during the loading process. This information can be crucial for assessing mechanical dynamics and identifying any potential differences in operator practices.

3.3.5. Bucket Angle Algorithm. The bucket angle algorithm function calculates the angular displacement of the bucket during the digging subcycle in the data file. It takes two inputs: **s** the $n \times 4$ matrix containing the row numbers in the sensor data that mark the start and end of the swing-in, digging, and swing-out sub-cycles (Figure 3.6). **BucketAngle** is the vector containing the bucket angle data from the data file. The output, **BuA**, is a vector that contains the angular displacement of the bucket during the digging subcycle for each cycle.

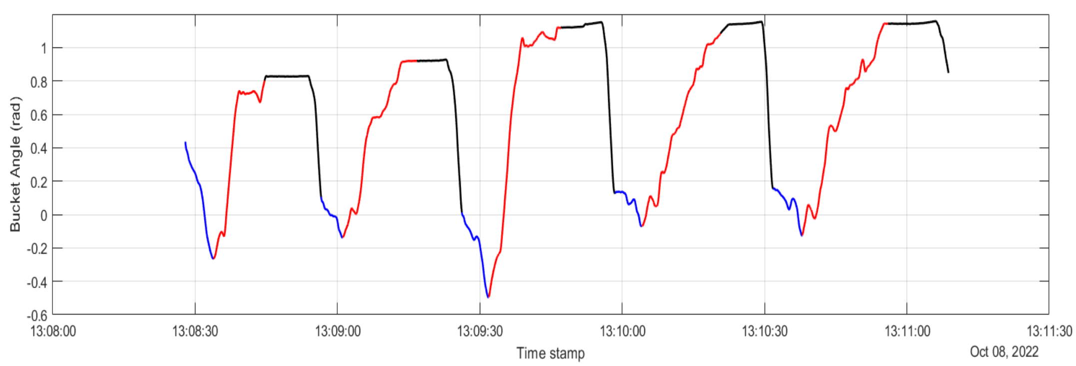

Figure 3.8 Structure of "typical" bucket angles during a cycle (blue = swing in; red = digging; and black = swing out subcycles)

For each digging subcycle (denoted by red color in Figure 3.8) defined by the row in **s**, the function identifies the BucketAngle at the start (second column of **s**) and the end (third column of **s**) of the subcycle (Figure 3.6). The function then calculates the angular displacement of the bucket for each digging subcycle, **BuA**, by subtracting the 'identified start' from the 'identified end.' By computing the angular displacement of the bucket for each digging subcycle, this function provides a key measure of the movement and operation of the bucket during the digging phase, offering insights into the performance and efficiency of the digging process. The author hypothesizes that it would also be helpful in explaining differences in operator practices.

3.3.6. Stick Angle Algorithm. The stick angle algorithm calculates the angular range of the stick during the digging subcycle in the sensor data file. It takes two inputs: **s** and **StickAngle**, where **s** is the $n \times 4$ matrix that contains the row numbers in the sensor data that mark the start and end of the swing-in, digging, and swing-out subcycles (Figure 3.6). **StickAngle** is a vector containing the stick angle data from the sensor data file. The output of the function is **StA**, a vector. It contains the range of stick angles for each digging subcycle.

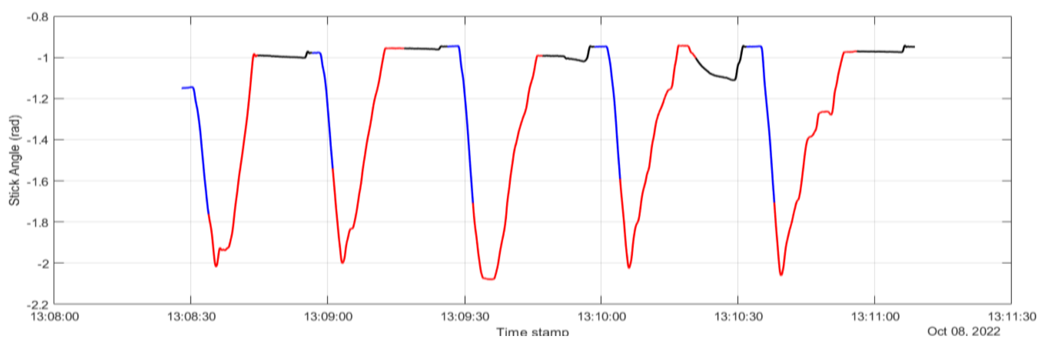

Figure 3.9 Structure of "typical" stick angles during a cycle (blue = swing in; red = digging; and black = swing out subcycles)

In each digging subcycle, defined by the row in **s** (Figure 3.6), the function identifies the segment of the StickAngle data corresponding to the digging phase (second and third columns of s). It then calculates the minimum and maximum values of StickAngle within this segment (Figure 3.9). The difference between the maximum and minimum angles gives the range of stick angles for that digging subcycle, which is stored in the corresponding index in StA.

By computing the range of stick angles for each digging subcycle, this function provides crucial information on the scope of stick movement during the digging process. This data can be essential for evaluating mechanical dynamics and possibly pinpointing differences in operator efficiency.

3.3.7. Swing Angle Algorithm. This function calculates each cycle's swing-in and swing-out angles from the sensor data file. This function takes two inputs: **s** and **SwingAngle**, where **s** is the $n \times 4$ matrix, and each of its rows contains the start and end indices of the swing-in, digging, and swing-out subcycles within the sensor data. **SwingAngle** is a vector that contains the swing angle data for each data point from the sensor data file. The function output consists of two vectors $n \times 1$, **SiA** and **SoA**, representing each cycle's swing-in and swing-out angles, respectively.

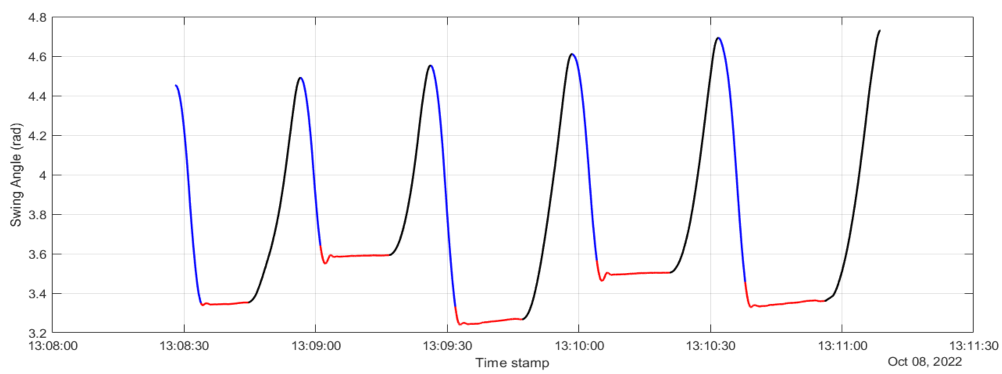

Figure 3.10 Structure of "typical" swing angles during a cycle (blue = swing in; red = digging; and black = swing out subcycles)

For the swing-in phase, the function identifies the initial and final **SwingAngle** values corresponding to the start (first column of **s**) and end (second column of **s**) indices of the swing-in subcycle (Figure 3.6). It then calculates the change in **SwingAngle** by subtracting the end angle from the start angle and stores this in **SiA**. The process is repeated for the swing-out phase, where the initial and final **SwingAngle** values correspond to the start (third column of **s**) and end (fourth column of **s**) indices of the swing-out subcycle (Figure 3.6). However, in this case, the change in **SwingAngle** is calculated by subtracting the start angle from the end angle, and this result is ultimately stored in **SoA**.

Figure 3.10 shows typical swing angle signals during a cycle. By determining the swing-in and swing-out angles for each cycle, this function provides valuable insight into the range of movement during the swing phases of operation, which can be critical for assessing operator efficiency and potential areas for improvement.

3.3.8. Dump Height Algorithm. The pre-processing algorithm provided by the commercial partner estimates the height of the bucket at any time step using the kinematics of the shovel. This data serves as the basis for estimating the dump height, another crucial performance indicator. This function calculates the vertical displacement of the bucket during the swing-out phase of each cycle based on the data from the sensor data file. The function takes in two inputs: **s** and **BucketHt**, where **s** is the $n \times 4$ matrix, with each row containing the start and end indices of the swing-in, digging, and swing-out subcycles within the sensor data (Figure 3.6). **BucketHt** is a vector that contains the bucket height data for each data point from the sensor data file (Figure 3.11). The output of the function is vector, **DH**, containing the vertical displacement of the bucket from the end of the digging phase to the end of the swing-out phase for each cycle.

The algorithm works by first identifying the bucket height at the end of the digging phase (indexed by the third column of **s**) and at the end of the swing-out phase (indexed by the fourth column of **s**). It then calculates the difference in bucket height between these two points, storing this vertical displacement in **DH**. By doing so, the dump height function provides a measure of the vertical distance the bucket travels during the swing-out phase, which can be useful in evaluating the effectiveness and efficiency of the dumping operation.

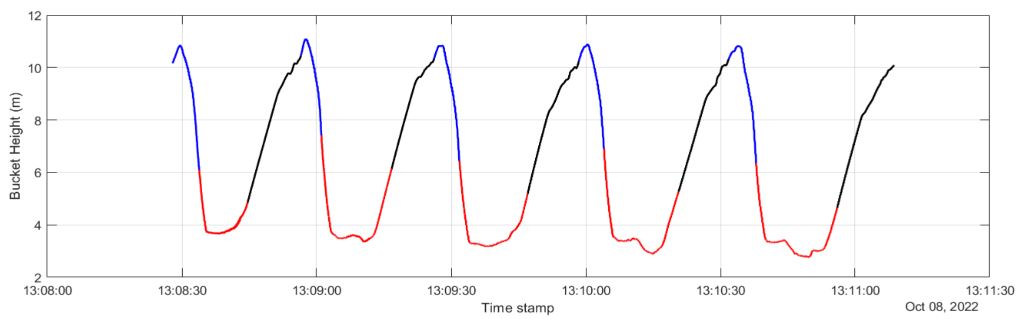

Figure 3.11 Structure of “typical” bucket heights during a cycle (blue = swing in; red = digging; and black = swing out subcycles)

3.3.9. Operator Identification Algorithm. The operator identification algorithm returns the identifier of the operator associated with each cycle of the sensor data. It accepts two inputs: **c** and **Operator**, where **c** is the $n \times 2$ matrix, with each row containing the start and end indices of a cycle in the sensor data (Figure 3.6). **Operator** is a vector containing the operator identifiers associated with the respective data points. The output of this function, **OP_id**, is a vector where each entry represents the operator identifier for the corresponding cycle.

The algorithm identifies the operator identifier at the start of each cycle (indexed by the first column of **c**) and assigns this value to **OP_id**. This means that the operator who initiated the cycle is considered the operator for the whole cycle. The operator identification function provides a simple way to track which operator was responsible for each cycle, facilitating analysis of operator-specific performance and efficiency.

3.3.10. Automation Algorithm. After developing the unique algorithms for extracting key performance indicators such as cycle time components, payload, energy, boom angle, bucket angle, dump height, operator ID, stick angle, and swing angles, an overarching function or algorithm was developed that brought all these individual pieces together, facilitating the automation of the entire process. This high-level algorithm acts as a comprehensive processing system for raw sensor files, running them all simultaneously through the established procedures. In doing so, it systematically extracts and collates the relevant information pertaining to the key performance indicators.

The automation function retrieves and processes sensor data from multiple CSV files, potentially across multiple folders, combining the data into a single data structure for further analysis. It allows the user to manually select folders, which are then processed one at a time. The function returns 15 different outputs: **OP_id** is an $n \times 1$ cell array that stores the operator identifiers associated with each cycle. **CT, SiCT, DiCT, SoCT, DH, FP, E, EPL, BA, BuA, StA, SiA,** and **SoA** are $n \times 1$ matrices representing different variables extracted from the data. These variables refer to cycle time, swing-in cycle time, digging cycle time, swing-out cycle time, dump height, final payload, boom's work (energy), energy per unit loading rate, boom angles, bucket angles, stick angles, swing-in angles, and swing-out angles. **VData** is an $n \times 13$ matrix that stores all the above variables except **OP_id** in one single data structure, where n is the total number of cycles across all processed files.

This function starts by initializing empty structures to hold the concatenated data and operator identifiers. It then prompts the user to select a directory. It iterates over all CSV files within the chosen directory, processing each file with the pre-processing algorithm provided by the commercial partner and extracting the operator identifiers and cycle variables. The extracted data is then appended to the previously initialized structures.

After all files in a directory are processed, the function prompts the user to choose whether to process another directory. If the user opts not to, the function ends the process, displays the concatenated data, and calls a separate function to extract individual variables from the concatenated data (i.e., to separate the variables into multiple vectors such that each variable stands alone as an individual vector) before finally returning all the outputs. This function is beneficial in cases where sensor data is split across multiple files or directories, allowing for efficient and streamlined processing of all available data.

3.4. SUMMARY

This research has successfully developed an algorithm that can sample “ideal” cycles (i.e., cycles with clear demarcation of stages, no bench clean up, and where the swing is in the direction where the operator’s vision is not unobstructed by the boom) from the monitoring file from the commercial hydraulic shovel monitoring system used in this research. Manual validation using visual inspections of plots of swing, boom, and stick angular displacements and the state enum variable to validate the cycle sampling algorithm show that the developed algorithm is 98% accurate.

In order to accomplish the second objective of this research, there is a need to develop algorithms to extract specific key performance indicators (KPIs) that help explain differences in operator energy per unit loading rate. Based on the nature of the monitoring data and a review of the literature focused on finding factors that are likely to explain differences in operator energy per unit loading rate, this work selected eight KPIs for each cycle: cycle time and cycle time components, payload, energy use, dump height, and boom, swing, and stick angles. Additionally, this work developed an algorithm to extract operator identities to facilitate separating data by operator. This work has successfully developed algorithms, implemented in MATLAB as MATLAB functions, to extract these KPIs from the shovel monitoring data.

4. PRELIMINARY DATA ANALYSIS OF PERFORMANCE DATA

The current section conducts preliminary statistical data analysis to understand the composition and characteristics of the performance data before any further investigation takes place. Another objective of the preliminary data analysis is to determine which operators from the complete list have enough data to be included in the analysis in subsequent sections. This section provides a preliminary investigation of data obtained from a hydraulic shovel with a bucket capacity of 40.5 yd^3, monitored via a commercial monitoring system. The data are visually and numerically examined using statistical and machine learning tools provided by MATLAB.

4.1. STRUCTURE AND STATISTICAL SUMMARY OF PERFORMANCE DATA

Following the successful development of the algorithms, they were deployed to extract the essential performance indicators (Sections 3.2 and 3.3) necessary for this study. The dataset under examination was retrieved from the monitoring database of a single hydraulic shovel in operation at the mine site. In a month, 1,809 cycles were sampled from the database, aided by the newly developed algorithms that helped transform the raw data into analyzable components. The algorithms extracted 12 distinctive parameters from these cycles, which depict the shovel's operational positions, time distribution across various cycles and sub-cycles, energy consumed by the boom, dump heights, and swing angles.

These parameters, extracted from the monitoring data in each cycle, form the heart of the analysis carried out in this research. They provide a detailed view of the various facets of the hydraulic shovel's operation, thereby enabling a comprehensive understanding of its energy consumption and productivity and how the operators' actions and behaviors influence them. Moreover, it is vital to recognize that the extracted KPIs or parameters fall into two distinct categories: linear and circular (directional). This separation is essential due to the distinct statistical analyses required for each data type (Berens, 2009; Zar, 1941). Linear parameters include cycle and sub-cycle times, payload, dump height, and energy. These parameters involve direct measurements with distinct start and end points, and their analysis usually employs traditional statistical techniques.

On the other hand, circular (directional) parameters, which include boom, bucket, stick, and swing angles, differ. These parameters are cyclical in nature, lacking clear start or end points as they circle back onto themselves. Hence, their statistical analysis necessitates specific tools, like those provided by circular statistics (Berens, 2009, 2023; Fisher, 1993; Zar, 1941), for proper interpretation and evaluation of the data.

A detailed statistical summary of these parameters, segregated into circular and linear, is presented in Table 4.1 and Table 4.2, respectively. These tables provide an overview of the vast amount of data extracted by the developed algorithms.

Table 4.1 Summary of relevant circular (directional) parameters (n = 1,809)

Parameters	Boom Angle (rad)	Bucket Angle (rad)	Stick Angle (rad)	Swing in Angle (rad)	Swing out Angle (rad)
Mean	0.6113	1.3073	0.9680	1.0454	1.2042
Median	0.6071	1.3305	0.9847	1.0339	1.2314
Deviation	0.0963	0.2899	0.1611	0.2584	0.4684
Variance	0.0046	0.0420	0.0130	0.0334	0.1097
Skewness	-0.0001	0.0195	0.0038	0.0008	0.0989
Kurtosis	0.9816	0.8488	0.9497	0.8806	0.7464

Table 4.2 Summary of relevant linear parameters (n = 1,809)

Parameters	Cycle time (secs)	Swing in time (secs)	Digging time (secs)	Swing out time (secs)	Dump height (m)	Payload (tonnes)	Boom Energy (kJ)
Minimum	12.8	1.7	2.7	2.7	-6.4	16.0	2685.2
Maximum	374.3	55.5	160.1	252.5	7.6	120.0	56125.1
Mean	31.9	5.5	16.1	10.2	3.6	54.8	15083.6
Median	29.9	5.3	14.9	9.4	3.7	46.8	14727.0
Deviation	13.1	2.0	7.5	7.0	1.5	23.0	3360.5
Variance	170.6	4.0	55.6	49.0	2.2	530.9	11292712.4
Skewness	16.0	11.5	9.5	25.3	-1.9	0.8	2.3

4.2. DETECTING AND REPLACING OUTLIERS

Outliers represent observations within a dataset that deviate significantly from other observations. They lie outside the overall pattern of distribution. These data points can potentially distort the interpretation and decrease the reliability of findings. Addressing outliers is crucial in data analysis for several reasons. Primarily, outliers can significantly influence the mean and standard deviation of the data, leading to inaccurate estimates. They can also distort the true underlying statistical relationships, cause overfitting in machine learning models and diminish the predictive performance of many machine learning algorithms.

However, the approach to handling outliers depends heavily on the characteristics and specific circumstances of the data. Complete removal of outliers could potentially lead to the loss of valuable information. This is especially valid in situations where the dataset is small, as removing outliers can further reduce the sample size and compromise the statistical power of the analysis. In the context of this research, due to the limited data points available, the author decided to replace the outliers rather than remove them.

Replacing outliers can be a preferred method to maintain sample size for statistical robustness. This process usually involves substituting the outlier with a measure such as the mean or median of the rest of the data or a predicted value from a suitable regression model. By doing so, the size of the dataset is preserved and, thus, prevents the loss of information that could have resulted from outright deletion while mitigating the negative impacts of outliers on the analysis.

Addressing outliers in this study involved distinct procedures for linear and circular data. The researcher used the Interquartile Range (IQR) method to identify outliers in linear data. This method is particularly effective for non-normally distributed linear data (NIST/SEMATECH, 2013). This method determines outliers as values that fall outside the range of median ± threshold, where the threshold is the interquartile range multiplied by a constant. This work set the threshold as 1.5 times the interquartile range. After identifying the outliers, they were substituted with the median value of the dataset. This approach ensures that the outliers in the linear data are effectively replaced, hence preserving the size of the data. Figure 4.2 and Figure 4.4 illustrate the data before and after replacing outliers.

In contrast, the work defined outliers for the circular data using a Z-score threshold. Here, the Z-scores represented the distance between each data point and the circular mean, normalized by the standard deviation. Any points with an absolute Z-score greater than the set threshold were classified as outliers. This Z-score threshold for this work is set at 3. The analysis replaced outliers with the circular median, a representative value for circular data. This method ensures an accurate identification and replacement of outliers within the circular data, thereby enhancing the overall quality and dependability of the data for further analysis. Figure 4.1 and Figure 4.3 shows the plot of the distribution of circular parameters before and after replacing outliers, respectively.

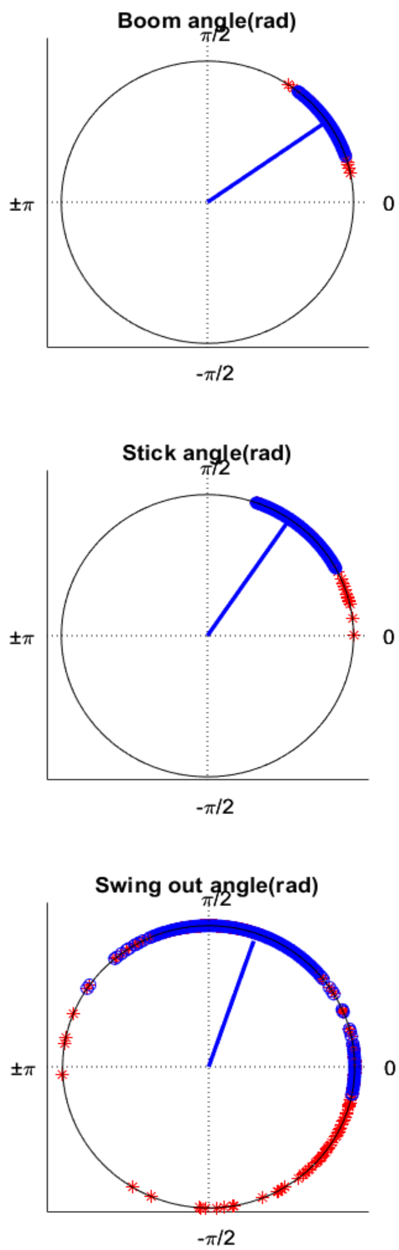

Figure 4.1 Polarized distribution plots of circular parameters prior to replacing outliers

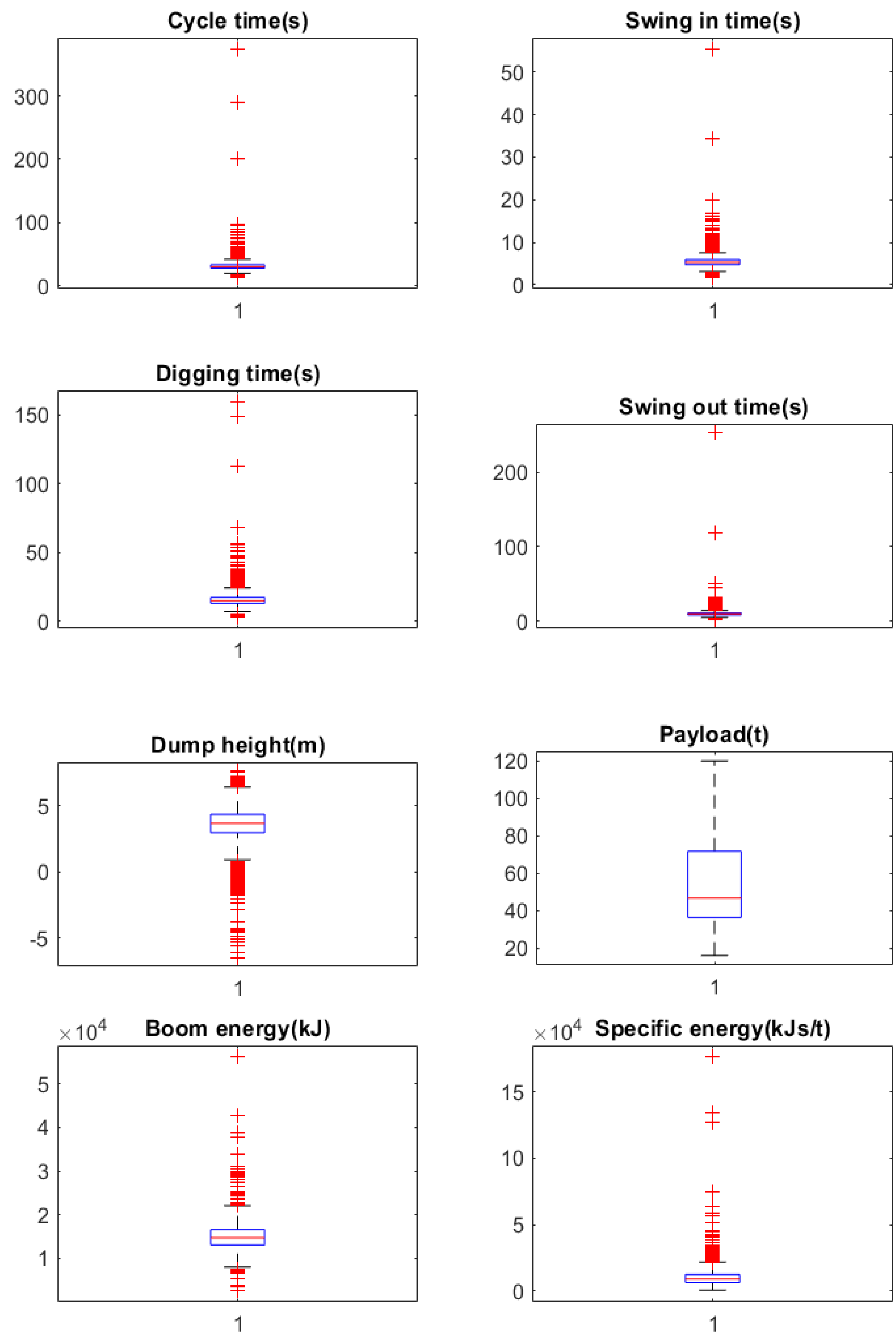

Figure 4.2 Boxplots of linear parameters prior to replacing outliers

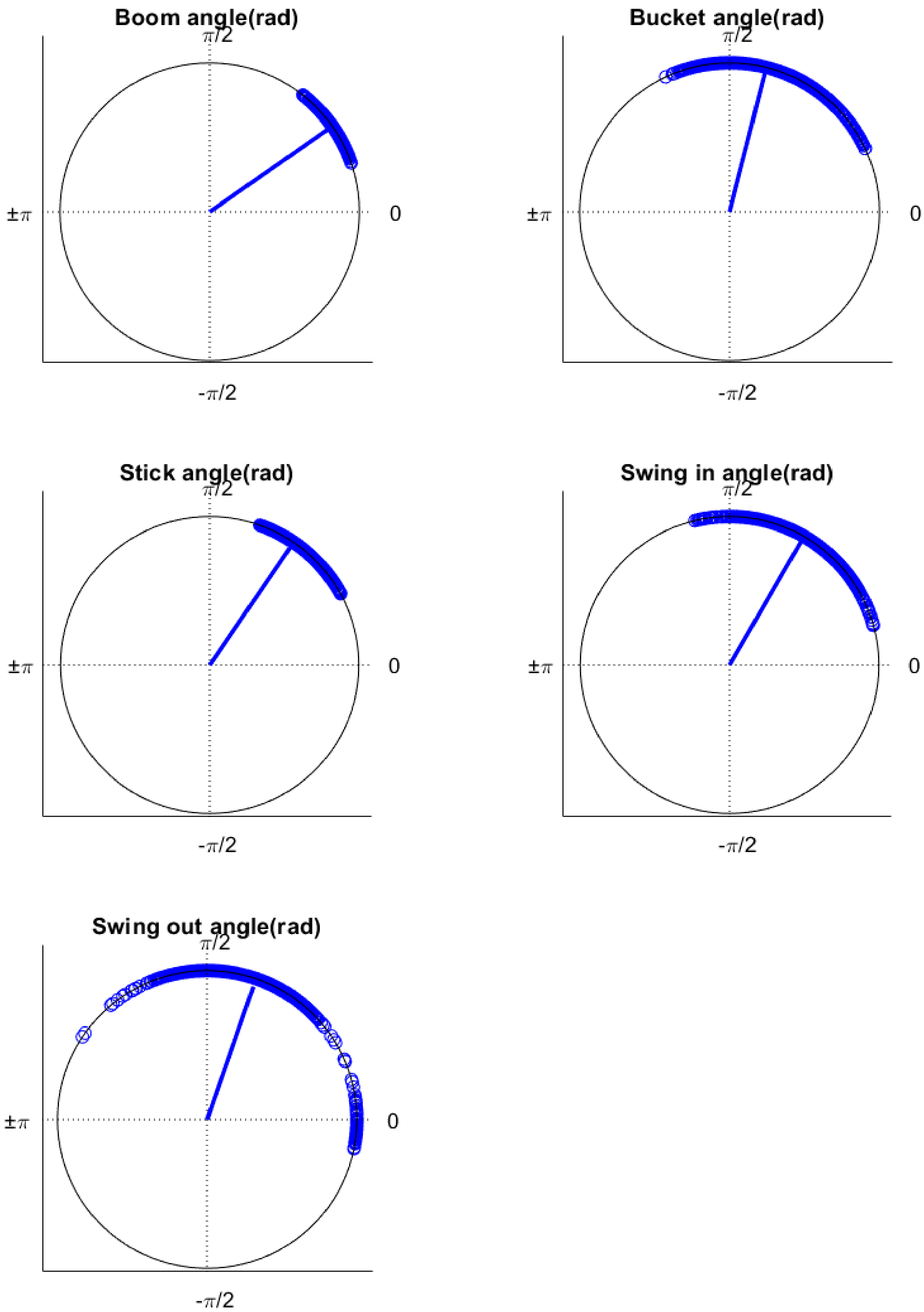

Figure 4.3 Polarized distribution plots of circular parameters post-replacing outliers

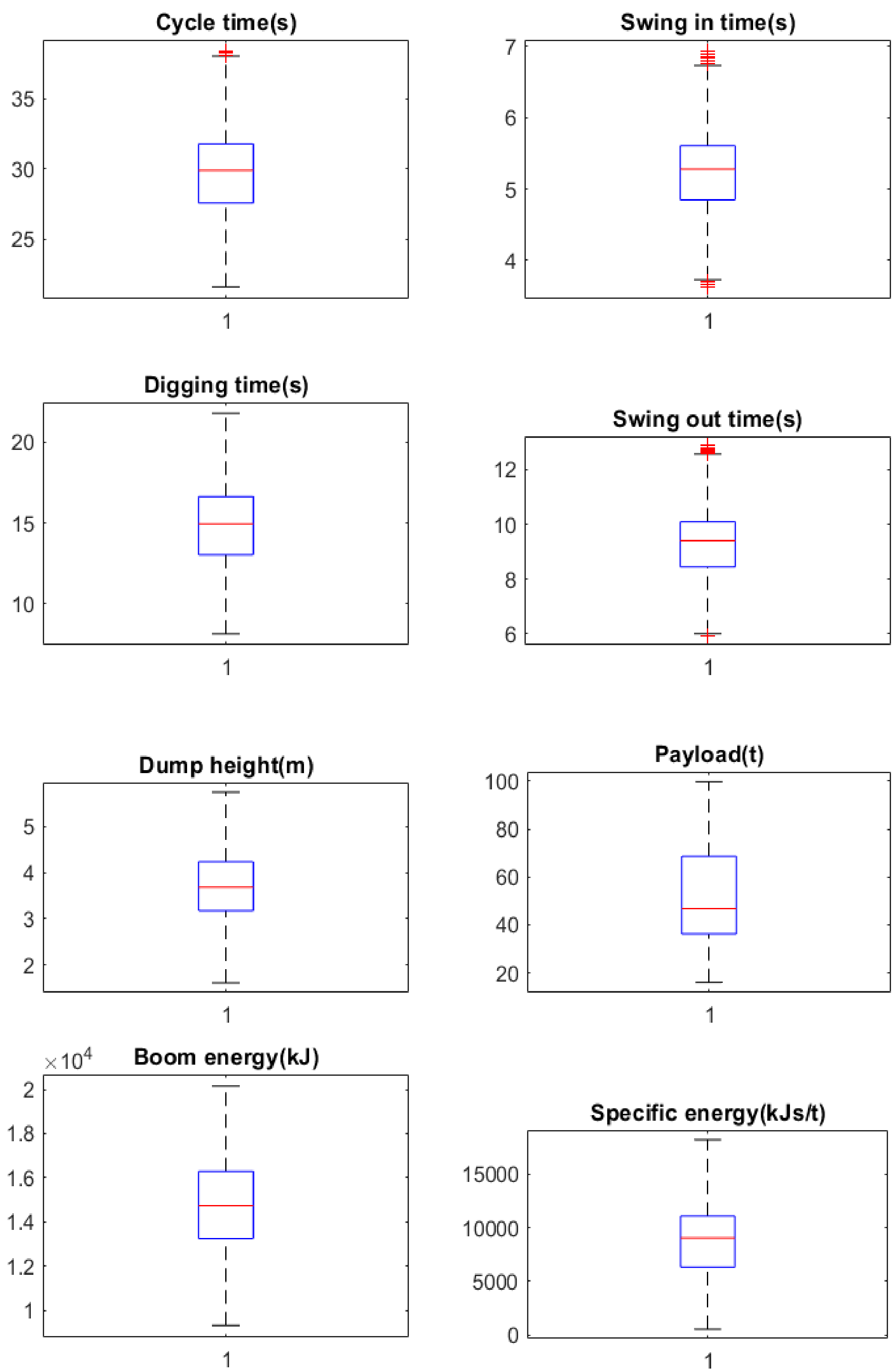

Figure 4.4 Boxplots of linear parameters post-replacing outliers

4.3. HYDRAULIC SHOVEL OPERATORS

In the quest to improve productivity and reduce energy consumption in mining, it becomes crucial to evaluate the performance of operators handling hydraulic shovels with a measure that considers productivity and energy consumption (Abdi-Oskouei, 2013). As discussed earlier in this work (Section 2), energy efficiency is a measure that combines energy consumption and useful work (production in the context of mining) into a single metric. A widely used metric for evaluating shovel energy efficiency is the energy per unit loading rate (Awuah-Offei & Frimpong, 2007; Patnayak, 2006), essentially the inverse of energy efficiency.

This researcher uses the energy per unit loading rate for this work (Equation 2.2) as a metric to evaluate the performance of operators. This metric is a good measure of digging performance (Awuah-Offei & Frimpong, 2007). Higher energy and cycle times will result in higher energy per unit loading rate, which is undesirable. So, reducing the energy per unit loading rate is optimal as the single parameter combines energy consumption and production rate in one measure.

$$\eta_l(i) = \frac{E_t(i)}{P(i)} = \frac{E_b.t_c(i)}{P(i)}\left(\frac{kJsec}{tonnes}\right) \tag{4.1}$$

In estimating the energy per unit loading rate, this researcher used the energy expended by the shovel’s boom in place of the overall energy consumption because the available monitoring data did not include information to estimate other forms of energy (swing energy in particular). Thus, Equation 4.1 represents the energy per unit loading rate in this work. The work done by the boom (boom's energy) is utilized instead of the total energy.

From the data collected over one month, 1,809 cycles are sampled using the algorithms developed in Section 3. The data included 15 unique operators; 1,584 out of the cycles belonged to the 15 operators, while no operators were identified for the remaining cycles due to gaps in the monitoring data, so they were discarded. Table 4.3 summarizes each operator's energy efficiency (energy per unit loading rate) derived from the gathered data.

Table 4.3 Summary of energy per unit loading rate of all operators

Operator	No of Cycle	Minimum (kJsec/tonnes)	Maximum (kJsec/tonnes)	Mean energy per unit loading rate (kJsec/tonnes)	Standard deviation of energy efficiency
A	319	1,371	16,290	8,679	2,902
B	256	3,086	18,809	10,309	3,346
C	168	3,163	21,653	10,108	4,274
D	153	1,938	17,364	8,161	3,811
E	146	2,101	22,690	10,610	4,617
F	129	2,007	15,597	8,090	3,229
G	119	3,756	15,339	7,855	2,542
H	73	3,399	11,847	7,930	2,215
I	63	2,854	13,146	7,708	2,701
J	37	2,474	16,310	8,708	2,872
K	35	2,975	11,529	6,435	2,068
L	34	3,882	15,783	8,554	2,964
M	24	3,038	16,267	8,583	3,529
N	16	3,609	13,714	7,996	3,301
O	12	5,263	13,584	8,822	2,685

4.4. HYDRAULIC SHOVEL OPERATOR SELECTION

It is crucial in the analytical process to ensure that every operator involved contributes substantially, allowing for credible deductions. It is not uncommon that some operators may not log enough hours to warrant inclusion in the analysis (Abdi-Oskouei, 2013). Mines often have varying operator schedules, particularly for their most significant and energy-demanding units. For example, a trainee operator might be allowed to operate a few hours per week under supervision, whereas the “main” production operators operate most of the time. These varying operator schedules can result in vast differences in the number of cycles each operator generates in the same period.

Further issues surface from standard workforce challenges, such as absenteeism and variation in operator skills in a shift-oriented work setting. These factors influence the total duration an operator spends handling a particular machine within a set time period (Abdi-Oskouei, 2013; Abdi-Oskouei & Awuah-Offei, 2016). Incorporating operators with insufficient work hours into the evaluation could potentially bias the outcomes, paving the way for incorrect inferences. An effective measure of the level of contribution each operator brings to the analysis can be represented by the mean standard error statistic of their energy per unit loading rate (Abdi-Oskouei, 2013; Abdi-Oskouei & Awuah-Offei, 2014; Biau, 2011).

$$SE = \frac{\sigma_{\eta}}{N_{cop}} \tag{4.2}$$

Standard error (SE) is a statistical term that represents the standard deviation of the sampling distribution of a sample statistic, most commonly the sample mean. In simpler terms, standard error is a measure of how spread out the means of different samples from the same population are likely to be, and it indicates the degree of precision of a sample statistic. It is important to note that the standard error is influenced by factors such as sample size and variability within the population. Generally, larger sample sizes lead to smaller standard errors, and more variable populations result in larger standard errors.

Table 4.4 Mean standard error of energy per unit loading rate of operators

Operator Id	No of cycles	Mean Standard error
A	319	9.10
B	256	13.07
C	168	25.44
D	153	24.91
E	146	31.62
F	129	25.03
G	119	21.36
H	73	30.35
I	63	42.87
J	37	77.62
K	35	59.09
L	34	87.18
M	24	147.06
N	16	206.30
O	12	223.75

Using Equation 4.2, the researcher estimated the mean standard error of energy per unit loading rate for each operator (Table 4.4). There is a noticeable increase in the slope of the mean standard error following Operator E, and both lines intercept on Operator I, as depicted in Figure 4.5. Any operators presenting a mean standard error exceeding that of Operator I (the intercept) are first eliminated from the datasets. The sharp rise in the slope of the mean standard error beyond Operator E indicates a shift in the standard error. Due to this abrupt change, the standard error of Operator E further serves as a reasonable cut-off value. So, this study uses Operators A to E (Figure 4.5) based on this analysis and informed engineering judgment to enhance operator count (i.e., the number of operators in the study) while preserving a reasonable confidence level in the mean energy efficiency estimates. Consequently, a mean standard error of 32 is the cut-off value used in determining the minimum required cycle number. Table 4.5 summarizes the overall performance of the operators selected for this study.

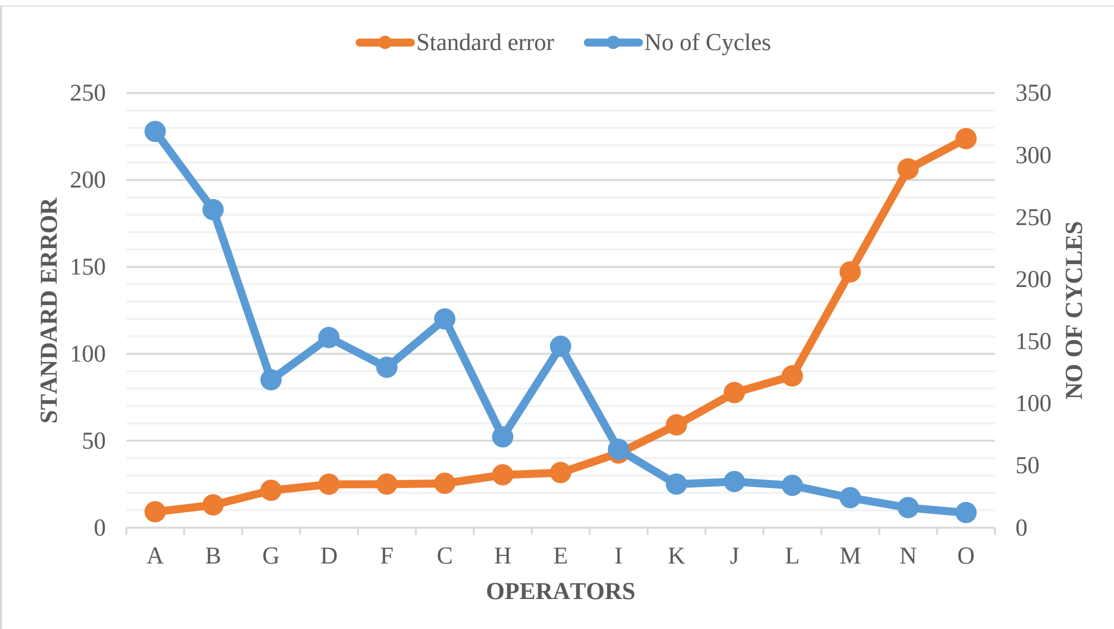

Figure 4.5 The mean standard errors plot against the cycles number of operators

Table 4.5 Summary of the overall performance of selected operators

Operator Id	No of cycles	Time (sec)	Energy consumption (kJ)	Production rate (tonnes/sec)	Mean Energy per unit loading rate (kJsec/tonne)
A	319	9,614.19	4,915,858	1.78	8,679
B	256	7,415.61	3,615,586	1.37	10,309
C	168	5,575.55	3,151,026	1.86	10,108
D	153	4,102.30	2,466,334	1.98	8,161
E	146	4,409.66	2,419,896	1.56	10,610
F	129	3,857.21	2,036,797	1.95	8,090
G	119	3,289.81	1,797,107	1.92	7,855
H	73	1,983.60	879,960	1.52	7,930

4.5. SUMMARY

This section conducted a preliminary data analysis of the performance data for the linear and circular data collected from a hydraulic shovel. This analysis included using the Interquartile Range (IQR) method for identifying outliers in the linear data and a Z-score-based approach for outliers in the circular data. These techniques effectively replaced outliers with median values without reducing the data sample size.

Furthermore, the section proposed an approach to selecting which operators should be included in the study. This study used the mean standard error statistic as the criterion with a cut-off value of 32 to decide which operators to include in the study. The researcher selected eight (8) operators to include in the study based on the mean standard error analysis.

5. EVALUATING THE INFLUENCE OF OPERATORS' PRACTICES ON HYDRAULIC SHOVEL ENERGY EFFICIENCY

This section explores the second objective of this research, which is to test the hypothesis that hydraulic shovel operator practices and skills affect shovel energy efficiency and identify critical parameters that explain the differences in operator energy efficiency. In an effort to present a detailed comparison, the researcher employs statistical evidence to illustrate the disparities in energy outputs among selected operators. The methodology for this examination is adopted from the work of Abdi-Oskouei (2013), providing a robust framework for this investigation. This approach is instrumental in facilitating analyses to accomplish the primary goal of this study. A significant segment of this section explores meticulous data analysis to identify the key parameters that are responsible for the differences in operators' energy efficiencies. This includes a correlation analysis between the practices of operators and their energy efficiencies and a difference regression analysis model.

5.1. ASSESSING DIFFERENCES IN OPERATOR ENERGY EFFICIENCY

In assessing the impact of operators' practices on the energy efficiency of hydraulic shovels, it is critical to employ a methodology robust enough to accommodate the significant variability observed in the performance metric (energy per unit loading rate) data (Ott & Longnecker, 2015; Ronald, 2016). This is evident from the initial analysis of the data from our case study, as shown in Table 4.1 and Table 4.2. Given the task at hand, it is important to substantiate any differences in the energy outputs among operators statistically. A mere comparison of the mean values of energy per unit loading rate would be insufficient and unreliable. This is because such an approach does not account for the possibility that observed differences could be attributed to chance, stemming from the specific samples taken. Therefore, it is essential to determine the significance of these differences statistically to ensure that they are not merely random fluctuations but represent real, meaningful variations in operator performance (i.e., differences in population mean). In statistics, t-tests and Analysis of Variance (ANOVA) are statistical tests commonly employed to compare means across distinct groups (Mishra et al., 2019). Each of these tests, however, is predicated upon specific assumptions. The ANOVA, for instance, assumes that:

- ✓ the observations from each group are normally distributed,
- ✓ the variances of all groups are equal (homogeneity of variance), and
- ✓ the observations are independent.

On the other hand, the t-test, used primarily for comparing two groups, assumes independent and identically distributed variables. In the context of this study, given that the researcher is comparing more than two groups (8 operators in this case), the researcher has chosen to apply the ANOVA test. It is imperative, though, to choose a statistical test appropriately, bearing in mind the nature of the datasets at hand. Each test has specific requirements or assumptions, and any violations could lead to misleading results or incorrect conclusions (Abdi-Oskouei, 2013; Herberich et al., 2010).

In addition, these violations may lead to either Type I or Type II errors. Type I error happens when a valid null hypothesis is incorrectly dismissed, resulting in a false positive. On the other hand, Type II error arises in scenarios where scientists erroneously accept an invalid null hypothesis, resulting in a false negative. Therefore, an incorrect choice or application of a statistical test could increase the chances of committing these errors, potentially leading to incorrect conclusions and decisions based on those conclusions. It is essential to acknowledge that alternative ANOVA methods are available for comparing means if datasets fail to meet any of the prerequisites or assumptions of traditional ANOVA. These alternatives can provide valid results under different conditions.

Initial data analysis is beneficial for enhancing data comprehension. Moreso, it aids in assessing whether the underlying assumptions of the tests are being met. In order to grasp the data, this researcher estimated the summary statistics for the eight operators under consideration. Table 5.1 shows these findings.

Table 5.1 Summary statistics of energy per unit loading rate of selected operators

Id	N	Minimum (kJsec/tonne)	Maximum (kJsec/tonne)	Mean (kJsec/tonne)	Standard deviation	Skewness
A	319	1,371	1,6290	8,679	2,902	0.4
B	256	3,086	1,8809	10,309	3,346	0.2
C	168	3,163	2,1653	10,108	4,274	0.5
D	153	1,938	1,7364	8,161	3,811	0.4
E	146	2,101	2,2690	10,610	4,617	0.6
F	129	2,007	1,5597	8,090	3,229	0.3
G	119	3,756	1,5339	7,855	2,542	0.4
H	73	3,399	1,1847	7,930	2,215	-0.5

The following sections delve into the analytical approach this study adopted for comparing means across the groups under consideration. Section 5.1.1 details the most suitable statistical tool in light of the outcomes of the verification process, while Section

5.1.2 details the steps to confirm whether the datasets satisfy the assumptions of ANOVA. Section 5.1.3 presents the results of applying this approach to the hydraulic shovel energy efficiency data in this work. Figure 5.1 shows a flowchart that illustrates selecting an appropriate method for comparing means.

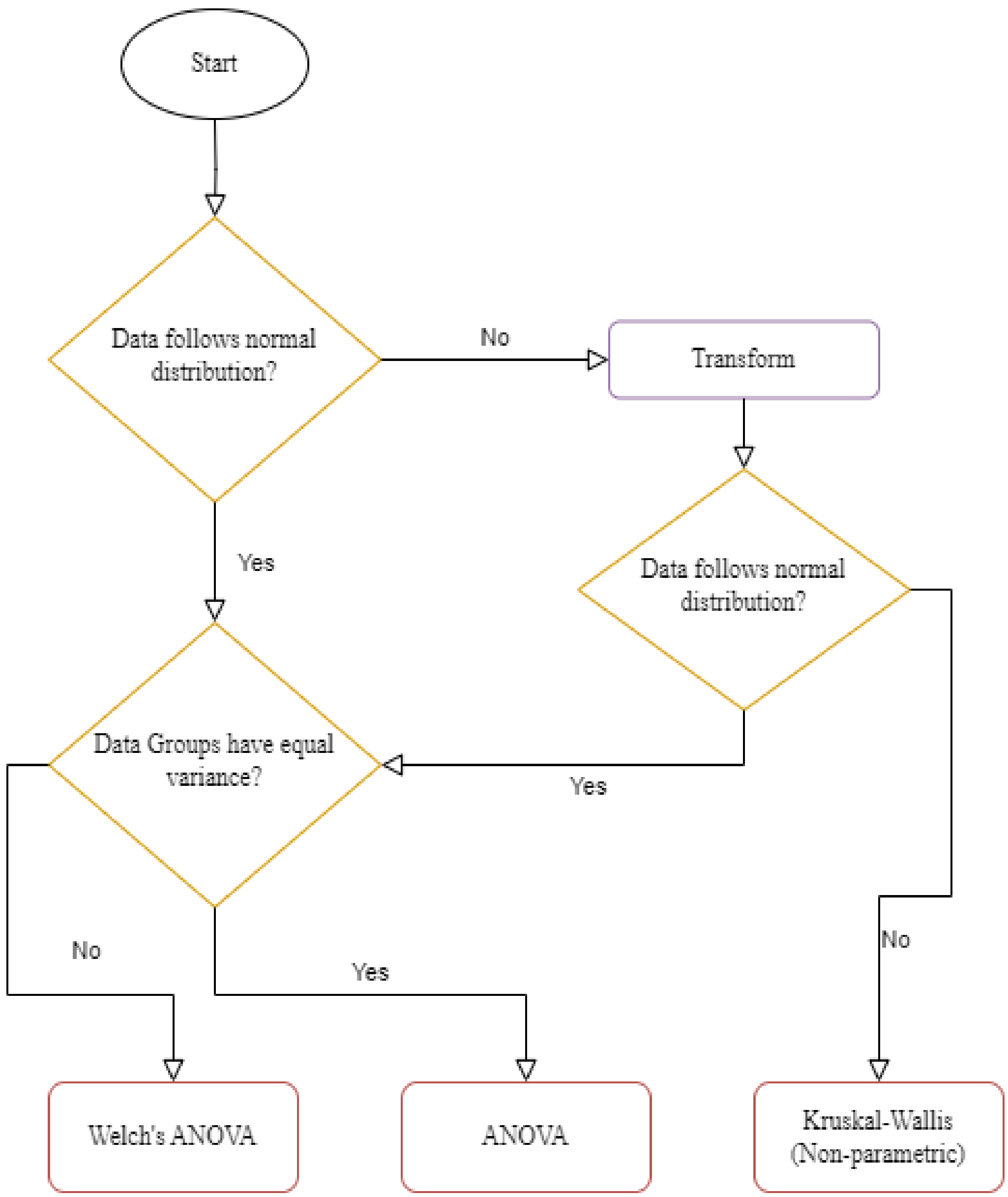

Figure 5.1 Flowchart illustrating the approach to comparing means (Abdi-Oskouei, 2013)

5.1.1. Equality of Means. The two prevalent techniques used in comparing means across different sample groups are the t-test and variance analysis (ANOVA). ANOVA examines the hypothesis of equal means across multiple groups (greater than two), with the null hypothesis that the means of all groups are equal. The t-test, on the other hand, compares the mean values when only two groups are involved in the analysis. Though straightforward, a t-test may lead to a Type 1 error (X. H. Zhou et al., 1997). This risk increases with multiple pairwise comparisons, as would be necessary when considering more than two operators in a dataset. Therefore, to mitigate the risk of Type 1 error, ANOVA is typically preferred when comparing more than two operators (Abdi-Oskouei, 2013).

However, datasets, such as the energy efficiency data in this study, may not always meet the normality's assumptions and homogeneity of variances crucial for ANOVA (Abdi-Oskouei, 2013). One approach to this problem is to change the form of the observations, often using a natural log transformation, but this approach can lead to its own set of issues. For example, the null hypothesis derived from log-transformed observation is not equivalent to the original hypothesis, particularly when the variances are unequal (X. H. Zhou et al., 1997). Consequently, there is a possibility of falsely rejecting the equality of the means in the initial observations, even after the null hypothesis of the transformed observations is not rejected. Therefore, any transformation should be used cautiously when the assumption of normality cannot be satisfied (Curran-Everett, 2017).

Welch's t-test and Welch's ANOVA (Welch, 1947), which relax the assumption of equal variances, can be effective solutions for situations where the assumption of equal variances is violated. These tests are practical, straightforward, and accurate with respect to Student's distribution, having the degrees of freedom determined by the magnitude of variance and the number of observations (Welch, 1947). Even when the variances are equal, Welch's t-test is frequently proposed as a preferable alternative to the traditional t-test (Krishnamoorthy et al., 2007; Rodgers & Nicewander, 1988).

In addition, non-parametric tests protect against the misapplication of statistical evaluations. Tests such as the Kruskal-Wallis (a non-parametric counterpart to ANOVA), along with the Wilcoxon-Mann-Whitney (an alternative to a t-test), carry fewer presumptions in comparison to parametric tests (Gibbons, 1985; Hollander & Wolfe, 2013). While this makes them somewhat less potent, it also reduces the likelihood of mistakes (Schlotzhauer, 2009). However, these tests may yield inaccurate results when applied to heavily skewed observations that have undergone log transformation (McElduff et al., 2010).

In summary, it is critical to validate underlying assumptions before using any of these statistical tests for testing equality of means. If the data is non-normal or does not meet the equality of variance assumption, the researcher should apply appropriate alternates to the ANOVA and t-tests.

5.1.2. Analysis of Variance. Analysis of Variance (ANOVA) is a statistical technique (parametric analysis) for comparing the mean values of multiple groups. It is based on the following three assumptions:

5.1.2.1. Independence of observations. This first assumption is more of a study design issue than something that can be tested with the data. It requires that the observations are collected independently of each other. The assumption of independence inherent in the ANOVA test appears plausible in the context of this study based on the belief that the energy efficiency demonstrated by one operator does not influence that of the other operators. Furthermore, it is essential to note that the data collection process was designed in such a way that an individual operator’s ID was utilized to obtain their respective energy output, reinforcing this sense of independence.

5.1.2.2. Normality. The responses for each group being compared are assumed to follow a normal distribution. This assumption does not necessarily mean that the combined distribution of all groups is normally distributed but that each individual group is. This assumption can be checked using normality tests (like the Shapiro-Wilk test) or by inspecting a histogram or a Q-Q plot. A combination of numerical and graphical methods can be employed to check if the observations (data) are normally distributed.

In this investigation, the researcher adopts both techniques to validate the normality of each operator’s energy efficiency data. Several tests can be applied to verify that the data (observations) are normally distributed, including the Kolmogorov-Smirnov (KS), Shapiro-Wilk (W), Cramer-von Mises (CVM), and the Anderson-Darling (AD) tests. The Shapiro-Wilk test, denoted as (W), is acknowledged as one of the most effective methods for checking normality in a data set. However, it is constrained by the size of the observation. Specifically, the sample size should range between 7 and 2,000 for this test to be valid (Shapiro et al., 1968; Stephens, 1974).

Tests like the Cramer-von Mises, Anderson-Darling, and Kolmogorov-Smirnov are suitable for massive observations. These tests operate on the principle of empirical cumulative distribution (Park, 2008; Schlotzhauer, 2009). If the null hypothesis of the Anderson-Darling test is rejected, it implies that the observations are not normally distributed when using the mean and variance of the observations. However, it could still be normal for other mean and variance values. The Cramer-von Mises and Kolmogorov-Smirnov tests also exhibit this shortcoming (Drezner et al., 2010; Stephens, 1974). Considering these limitations, it becomes evident that a balanced approach employing graphical and numerical techniques is most beneficial when examining data for normality.

For the specific case of this work, the researcher chose the Shapiro-Wilk (W) test as the preferred method to test for normality, as the datasets fall within the limit of 2,000 samples required for this test. Adopting the Shapiro-Wilk test ensured a robust and accurate examination of the normality assumption, allowing for more reliable interpretations and conclusions. Additionally, this comprehensive approach to normality testing, which integrates both numerical and graphical methods, enhances the overall reliability and validity of the data analysis process.

5.1.2.3. Homogeneity of variance. The ANOVA test assumes that all data groups should have equal variances, also known as homoscedasticity. F-test, Levene's, and Bartlett's tests are commonly used to verify this assumption between two or more groups of observations.

Bartlett's and F-tests are known to be highly sensitive to the normality of observations (Schultz, 1985). In contrast, Levene's test, first proposed by Levene (1960) and subsequently enhanced by Van Valen (1978, 2005), is a robust alternative to the F-test. It maintains its reliability even in instances where the observations do not follow a normal distribution (Levene, 1960; Van Valen, 1978, 2005). As such, for the purpose of this research, this researcher adopts Levene's test to check if the variances of the groups are equal.

5.1.3. Hydraulic Shovel Operator Energy Efficiency. Table 5.1 presents the summary statistics of energy per unit loading rate of selected operators. In an effort to determine whether there are statistically significant differences in the operator energy efficiencies, the researcher tested the equality of means across different groups (i.e., the energy efficiency of operators A to H) using the ANOVA test. This is crucial for proving that there is indeed a significant difference in the energy outputs of the operators. However, before implementing ANOVA, it was essential to confirm that the data conformed to its underlying presumptions - independence of observations, normality of the distribution, and homogeneity of variances.

This researcher first confirmed the independence of groups, a vital prerequisite for conducting an ANOVA. This ensured that the results in one group did not influence those of another. Following the verification of group independence, the next step was to test the normality of the data. The Shapiro-Wilk and Shapiro-Francia's normality test is employed to run the analysis. This test is implemented through the MATLAB function provided by BenSaida (2009), a robust statistical method designed to test the composite normality of a dataset ranging between 3 and 5000. This test examines the null hypothesis (H_0) that the population from which a random sample X is drawn is normally distributed with an unspecified mean and variance. This test is omnibus and is considered to be highly powerful against various alternatives. It distinguishes between types of non-normality, performing the Shapiro-Francia's test for leptokurtic (heavy-tailed or profusion of outliers) samples and the Shapiro-Wilk's test for platykurtic (light-tailed or lack of outliers) samples. Table 5.2 shows the results of the test.

Table 5.2 Normality tests result for the energy per unit loading rate of operators

Energy per unit loading rate	Opr A		Opr B		Opr C		Opr D	
Shapiro-Wilk Test	W	p	W	p	W	p	W	p
	0.976	0.000	0.988	0.028	0.961	0.000	0.963	0.000
	Opr E		Opr F		Opr G		Opr H	
Shapiro-Wilk Test	W	p	W	p	W	p	W	p
	0.948	0.000	0.976	0.024	0.967	0.005	0.929	0.001
Log (energy per unit loading rate)	Opr A		Opr B		Opr C		Opr D	
Shapiro-Wilk Test	W	p	W	p	W	p	W	p
	0.969	0.000	0.965	0.000	0.975	0.004	0.961	0.000
	Opr E		Opr F		Opr G		Opr H	
Shapiro-Wilk Test	W	p	W	p	W	p	W	p
	0.971	0.004	0.968	0.003	0.975	0.028	0.869	0.000

The results from the Shapiro-Wilk test (Table 5.2) indicate that the energy per unit loading rate of all the operators significantly deviates from normality ($p \leq 0.028$). This suggests that the null hypothesis of normality was rejected for all the operators at the 95% confidence, indicating that the data do not conform to a normal distribution. This result is the same after log-transforming the data ($p \leq 0.004$).

In this case, the log transformation did not successfully normalize the data. This result is not uncommon, as not all data can be effectively normalized using a log transformation. The researcher adopted additional tactics to examine the data in light of the statistical results, which consistently indicated a significant deviation from normality despite log transformation. Recognizing that statistical tests, despite their utility, can have limitations and might not always provide the complete picture, the researcher utilized graphical analysis as a supplementary tool. The researcher generated histograms and Q-Q pots to inspect the distribution of the observations visually. Visual analysis can often reveal patterns or characteristics that might not be immediately apparent through numerical summaries alone. Such plots can provide insightful visual cues about the skewness, kurtosis, and overall symmetry of the data (or observations) distribution.

After the log transformation, the graphical plots indeed showed a shift towards normality. Although the data still did not perfectly fit the assumption of normality according to Shapiro-Wilk’s test, the graphical visualization of the data demonstrated improvements. The data visualization supports the idea that the log transformation had some effect in pushing the data toward a more normal distribution. This indicates that the researcher’s decision to conduct a log transformation was not without merit, as it induced a degree of normality in the data even though it did not entirely satisfy the normality assumption. All the figures for the graphical method are in APPENDIX A.

The last assumption required to satisfy the conditions for Analysis of Variance (ANOVA) is the homogeneity of variance. In order to check this, the researcher employed Levene’s test for homogeneity of variances, using a MATLAB script provided by Trujillo-Ortiz and Hernandez-Walls (2003). Levene’s test assesses the null hypothesis that the variance is the same across all groups. In this test, the data are transformed such that Y_{ij} = abs[X_{ij} - mean(X_j)], and then a one-way ANOVA is performed using Y as the dependent variable. Essentially, this test checks whether the absolute deviations from the group means are equally spread across the groups. Figure 5.2 shows the results from the MATLAB script used in this work.

```
The number of samples are: 8

-----------------------------
 Sample    Size      Variance
-----------------------------
   1       319         0.1281
   2       256         0.1273
   3       168         0.1993
   4       153         0.2709
   5       146         0.2113
   6       129         0.1951
   7       119         0.1101
   8       73         0.1073
-----------------------------

Levene's Test for Equality of Variances F=11.4376, df1= 7, df2=1355
Probability associated to the F statistic = 0.0000
The associated probability for the F test is smaller than 0.05
So, the assumption of homoscedasticity was not met.

P =

   3.6415e-14
```

Figure 5.2 Results from Levene's test for equality of variance

The output for Levene's test on the log-transformed data (observations) for the operators shows that the assumption of homogeneity of variances was not met (Figure 5.2). This is evident from the associated p-value of the F-statistic, which is effectively 0. The F-statistic for Levene's test was 11.4376, with degrees of freedom 1 = 7 and degrees of freedom 2 = 1355 (Figure 5.2). The p-value associated with this F statistic is very small (approximately 0.0000), indicating strong evidence against the null hypothesis of equal variances across groups.

This suggests that the variance of the log-transformed data differs significantly among the operators. Therefore, this data does not satisfy the assumption of homogeneity of variance, which is critical for the validity of ANOVA. The researcher then considered alternate statistical methods which did not require this assumption or an investigation of the reasons behind the unequal variances and addressed these before proceeding with the ANOVA. Neither the original data nor its logarithmic transformation could meet these requirements for the ANOVA test. Although the log-transformed data came close to satisfying the normality assumption, it was still insufficient.

Given this scenario, the researcher utilized Welch's ANOVA (a variant that does not strictly require equal variances) on the log-transformed data. To avoid drawing potentially biased conclusions from inappropriate usage of statistical tests, the researcher also performed a non-parametric counterpart to ANOVA, the Kruskal-Wallis ANOVA test, on the original data, allowing a comparison of results.

Welch's ANOVA method works by comparing the test statistic to the F-distribution. It takes into account the size, mean, and variance of each group in the sample. This test is beneficial when the population variances are unknown or unequal. On the other hand, the Kruskal-Wallis test is a rank-based non-parametric test that can be used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable. It is essentially a nonparametric version of ANOVA. Table 5.3 shows the results of Welch's ANOVA, while Table 5.4 shows the results of the Kruskal-Wallis ANOVA test.

Table 5.3 Results of Welch's ANOVA test

Source	Degree of freedom	F-statistic	P-value
Treatment	7	15.7210	0.0000
Error	459.0369		

Welch's ANOVA test was carried out using a MATLAB script provided by Trujillo-Ortiz (2012). The results (Table 5.3) revealed a significant effect, as indicated by an F-statistic of 15.721 and a p-value of 0.0000. This means that the null hypothesis of equal sample means was rejected because the extreme test statistic suggests that at least one of the means differs from the others. The Kruskal-Wallis test resulted in a p-value that is effectively zero (Table 5.4). As a result, the null hypothesis of equality among the population medians is rejected because of the extreme test statistic. This also indicates significant differences between the groups in the non-parametric context.

Table 5.4 Result of the Kruskal-Wallis ANOVA test

Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Groups	1.54E+07	7	2.20E+06	99.34	1.48E-18
Error	1.96E+08	1355	1.44E+05		
Total	2.11E+08	1362			

In summary, Welch's ANOVA and Kruskal-Wallis tests suggest significant differences exist among the energy efficiencies of the eight operators. This researcher believes this is a reliable result because careful analysis supports the use of these two particular tests. In addition, the fact that the conclusions of the two tests mirror each other confirms the results. This is an important step in this research because it demonstrates a statistically significant difference between the performance (as measured by energy per unit loading rate) of the eight operators in the study. Without this conclusion, further work to explore why the differences in energy efficiency will be irrelevant.

5.2. IDENTIFYING THE KEY PARAMETERS THAT DIFFERENTIATE OPERATORS

Based on the statistically proven differences in operators' energy efficiencies, this study concludes that variations exist in the energy outputs among operators. Therefore, the next imperative task is to identify the parameters accountable for these differences in order to accomplish the second objective of this thesis (to test the hypothesis that hydraulic shovel operator practices and skills affect shovel energy efficiency and identify critical parameters that explain the differences in operator energy efficiency). However, this study first examines potential correlations with energy efficiency among the parameters extracted in Section 3 to identify these key parameters. Correlation analysis provides insights into the relationships between energy efficiency and these parameters. While this researcher recognizes that correlation does not imply causation, the general premise here is to use correlation analysis to find variables that are likely to influence energy efficiency (i.e., operator practices and skills that affect energy efficiency) and use these variables in "difference regression analysis" to determine which variables explain differences between operator energy efficiency.

This research employs a difference regression analysis to evaluate variables explaining operator energy efficiency differences. The analysis aims to pinpoint the parameters responsible for the observed differences in energy efficiency among the operators. Regression analysis will quantify how the dependent variable (energy per unit loading rate) changes in relation to one or more independent variables (parameters identified from the correlation analysis). This process will enable the quantification of the effect each parameter has on energy efficiency, aiding in the identification of the parameters significantly impacting energy efficiency.

5.2.1. Correlation Analysis. The parameters extracted in Section 3 comprise both linear and circular data types, necessitating different approaches for correlation testing with energy efficiency. Furthermore, the data sets for these observations do not follow a normal distribution. This is verified for the linear data using the Shapiro-Wilk test (Table 5.5). The Von Mises distribution and Q-Q plots established non-normality for the circular using data as shown in Figure 5.3 and Figure 5.4 (Berens, 2009; Fisher, 1993; Jammalamadaka & Sengupta, 2001; Stephens, 1969).

Given this, non-parametric methods were employed for correlation testing of the linear parameters. This study specifically utilized Spearman's correlation (Best & Roberts, 1975; Gibbons, 1985; Hollander & Wolfe, 2013). This non-parametric test measures the strength and direction of monotonic relationships between variables, providing an alternative to Pearson's correlation (a parametric test commonly used for correlation analysis). It takes values between -1 and 1. On the other hand, correlation testing for the circular parameters was more complex. A combination of Spearman's correlation and a "wrapToPi" function is utilized.

Table 5.5 Shapiro Wilk's normality test results for linear parameters

Parameters	H	SW Statistic	p-value
Cycle time	1	0.9848	<0.001
Swing in time	1	0.9882	<0.001
Digging time	1	0.9877	<0.001
Swing out time	1	0.9855	<0.001
Dump height	1	0.9922	<0.001
Payload	1	0.9117	<0.001

In circular statistics, data is typically assumed to be in the range $[0, 2\pi)$ or $[-\pi, \pi)$ because angles are periodic with a period of 2π. This means that an angle of 0 is equivalent to an angle of 2π, an angle of π is equivalent to an angle of $-\pi$, and so on. Therefore, when working with angular or circular data, it is a common practice first to normalize or "wrap" the data into a consistent range. The 'wrapTo2Pi' function in MATLAB is used to wrap angles in radians to the interval $[0, 2\pi)$. Similarly, wrapToPi wraps angles to the interval $[-\pi, \pi)$.

This approach is analogous to Spearman's rank correlation and aligns with the methods suggested for testing non-parametric linear-angular relationships (Fisher, 1993; Fisher & Lee, 1981; Mardia, 1976; Mardia & Jupp, 1999; Zar, 1941). The correlation results are presented in Table 5.6. The correlation results suggest that all parameters, except for the 'Bucket angle' and 'Swing out angle,' significantly correlate to energy efficiency ($p < 0.05$). 'Payload' shows the strongest negative correlation with energy efficiency, indicating that as payload increases, energy efficiency decreases. This is an important finding, as it suggests that adjusting the payload could have a significant impact on energy efficiency.

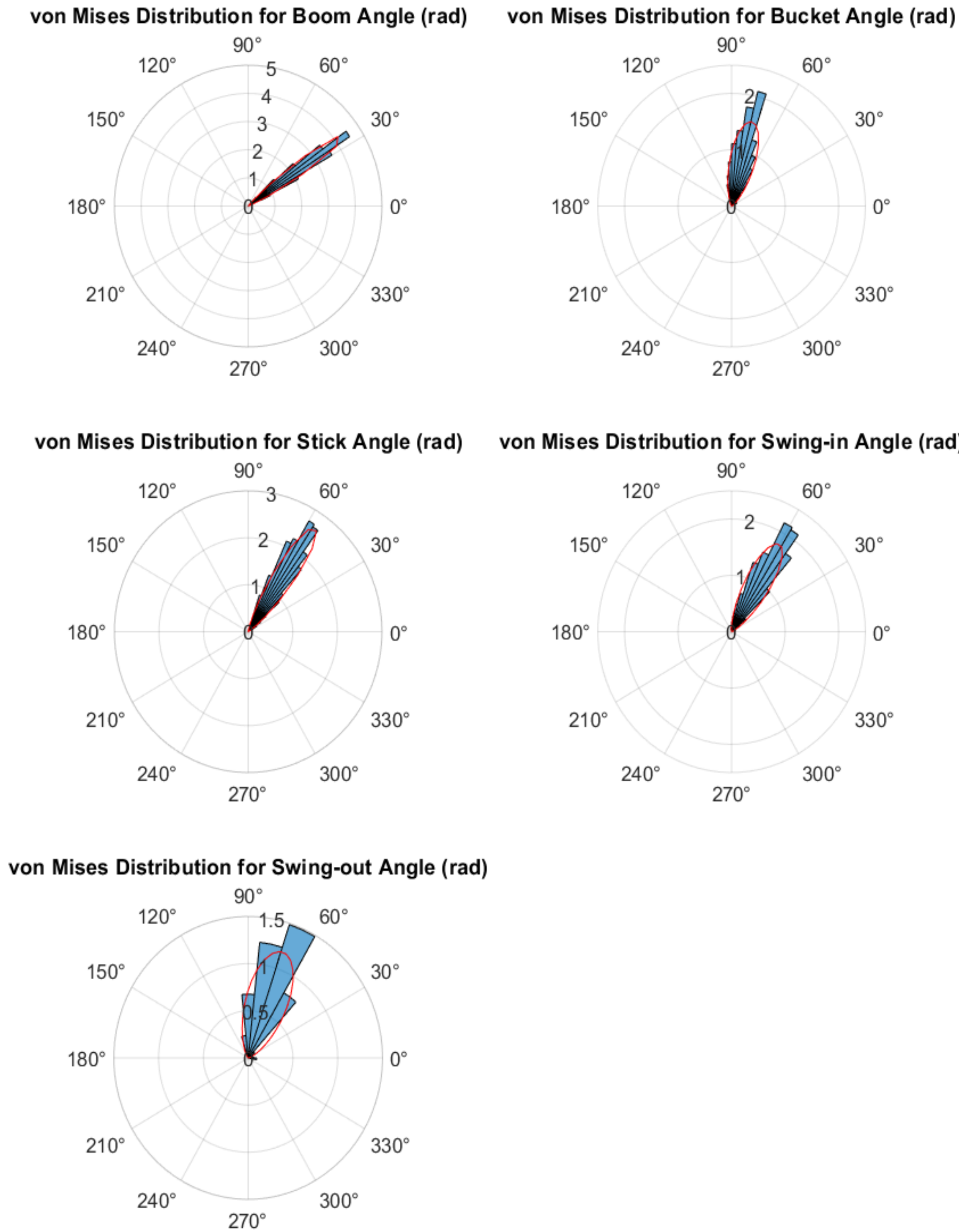

Figure 5.3 von Mises distribution plots of the circular parameters

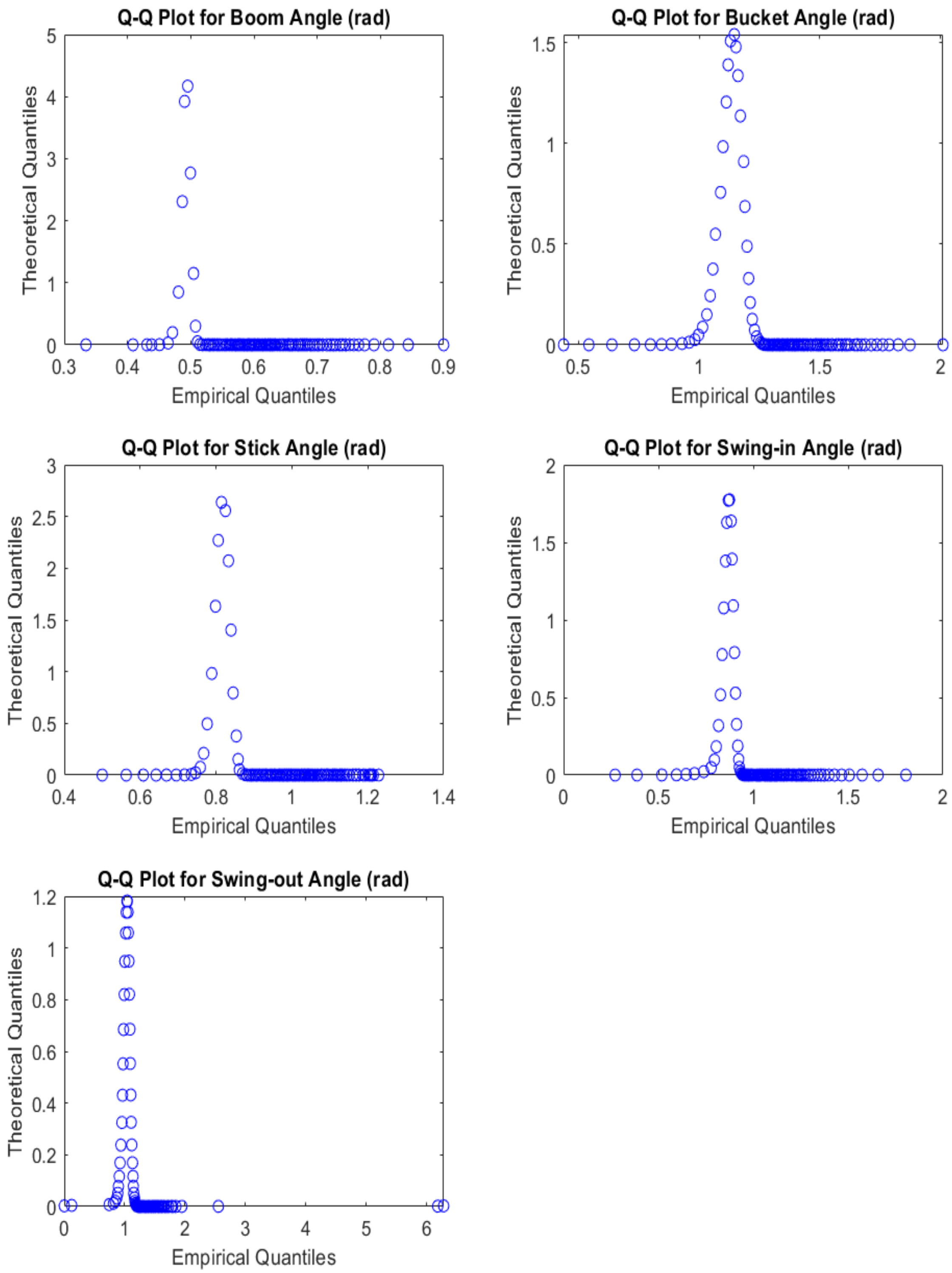

Figure 5.4 Q-Q plots of the circular parameters

Table 5.6 Results of correlation analysis

S/N	Parameters	Coefficient	p-value
1	Cycle time	0.2749	0.0000
2	Swing in time	0.0547	0.0199
3	Digging time	0.2282	0.0000
4	Swing out time	0.0961	0.0000
5	Dump height	0.0682	0.0037
6	Payload	-0.6472	0.0000
7	Boom angle	0.2588	0.0000
8	Bucket angle	0.0373	0.1132
9	Stick angle	0.1302	0.0000
10	Swing in angle	0.0477	0.0426
11	Swing out angle	-0.0015	0.9482

Interestingly, ‘Cycle time,’ ‘Digging time,’ and ‘Boom angle’ also show relatively strong positive correlations with energy efficiency. This implies that as these parameters increase, so does energy efficiency. Furthermore, the result suggests that modifications to these operational aspects can enhance energy efficiency.

5.2.2. Regression Analysis. Drawing from the work of Abdi-Oskouei (2013; 2016), this study leverages a difference linear regression model to identify the key parameters among the correlated ones that drive the disparities in energy efficiencies across different operators. Abdi-Oskouei’s model allows for a pair-wise comparison amongst a set of operators. For example, if there are an ‘*n*’ operators, the number of pairs of operators required for comparison can be evaluated using the combination formula provided in Equation 5.1.

$$^{n}C_{k} = \frac{n!}{k!(n-k)!} \tag{5.1}$$

Abdi-Oskouei (2013) did not discuss or specifically address circular variables in her difference regression, even though her variables included one circular variable (swing-out angle). This study, however, carefully evaluates the theory and basis for using circular variables in regression (Mohammad et al., 2021). Most of the literature on circular regression addresses circular-linear regression (Mohammad et al., 2021), where the response variable is circular, and the predictor variables are linear. There are a few examples of circular-circular regression models too (Kato et al., 2008). However, there are few examples of linear-circular regression models, and the existing models are complicated (Bhattacharya & SenGupta, 2009; Kim & SenGupta, 2015). In this work, the response variable (differences in energy per unit loading rate) is linear, whereas some of the predictor variables (differences in boom, swing-in, and stick angles) are circular. Therefore, the well-developed circular-linear regression models will not apply here.

Esmaieeli-Sikaroudi (2017) shows that linear-circular parameters could be modeled linearly under the condition that the data (observations) distribution for the circular parameters does not cross the $\mathbf{0}$ and $\boldsymbol{\pm\pi}$ boundaries. The distributions of the correlated circular parameters conform to this requirement, with no observations crossing these boundaries (as shown in Figure 4.3). Consequently, this study was able to model the difference regression linearly. Also, when dealing with circular variables, one obtains better results if there are attempts to “linearize” the variables (Kim & SenGupta, 2018; Pewsey et al., 2013). Examples include using the sine or cosine of the angle or a function that converts angles to a range of $-\pi$ to $+\pi$.

The approach adopted in this study is essentially a combination of relying on Esmaieeli-Sikaroudi’s finding of the performance of linear regression models under certain conditions and converting all angular data to a range of -π to +π before regression. This work successfully uses this difference linear regression approach to achieve the second objective of this study. This researcher believes this approach is an improvement upon Abdi Oskouei (2013) as it accounts for the circular variables in a way that she did not do. The developed difference regression model for hydraulic shovels (based on the monitoring system data this work uses) is presented as Equation 5.2 in this study. This thesis notes that the model for a hydraulic shovel based on a different monitoring system will likely result in a different set of variables. However, the approach in this work will be useful for identifying explanatory variables (given the available data) for the differences in operator energy efficiency to be helpful regardless of the set of variables.

$$\Delta\eta_l = k_0 + k_1\Delta C_t + k_2\Delta Si_t + k_3\Delta Di_t + k_4\Delta So_t + k_5\Delta D_h + k_6\Delta P + k_7\Delta B_a + k_8\Delta St_a + k_9 Si_a \tag{5.2}$$

Linear regression is a statistical method used to model the relationship between a dependent variable (also known as an outcome variable) and one or more independent variables (also known as predictors). In the context of this research, the dependent variable under investigation is the difference in energy per unit loading rate of the two operators under consideration. The independent variables, also known as predictors, are the differences between variables (or parameters) that were found to be significantly correlated with energy per unit loading rate in the prior analysis between the two operators under consideration. It is important to note that these predictors include both linear and circular data types, which are treated appropriately for the analysis.

The algorithm in this work relies on a difference matrix (by concatenating the dependent and predictor variables into a matrix) for pair of operators. The construction of the difference matrix necessitates having an equal number of cycles for each operator. However, achieving this in practice is rather challenging due to the inherent variability in cycle times. Even under circumstances where operators are allocated equal working hours, discrepancies in the number of cycles can arise. Thus, a random sampling approach is implemented to mitigate this issue and ensure robust analysis. This method allows for a fair comparison by matching pairs of operators who do not have an equal number of cycles.

Two algorithms implemented within MATLAB as functions are developed for this analysis. The first (base function) is designed to facilitate comparing two datasets by applying linear regression to the difference between operators' parameters. A series of linear regressions are carried out as part of this function's operations to determine the regression coefficients and test whether they are significantly different from zero at 95% confidence (by evaluating whether the 95% confidence intervals for the regression coefficients include zero).

A flowchart of the developed base function algorithm, provided in Figure 5.5, offers a clear, step-by-step depiction of how the process works. The base function compares two operators (OprA and OprB) using multiple linear regression analyses. The comparison involves iteratively running a linear regression on randomly sampled subsets of the datasets and checking if the confidence intervals of the regression coefficients span zero. This process repeats for *'i'* iterations to avoid bias in the sampling.

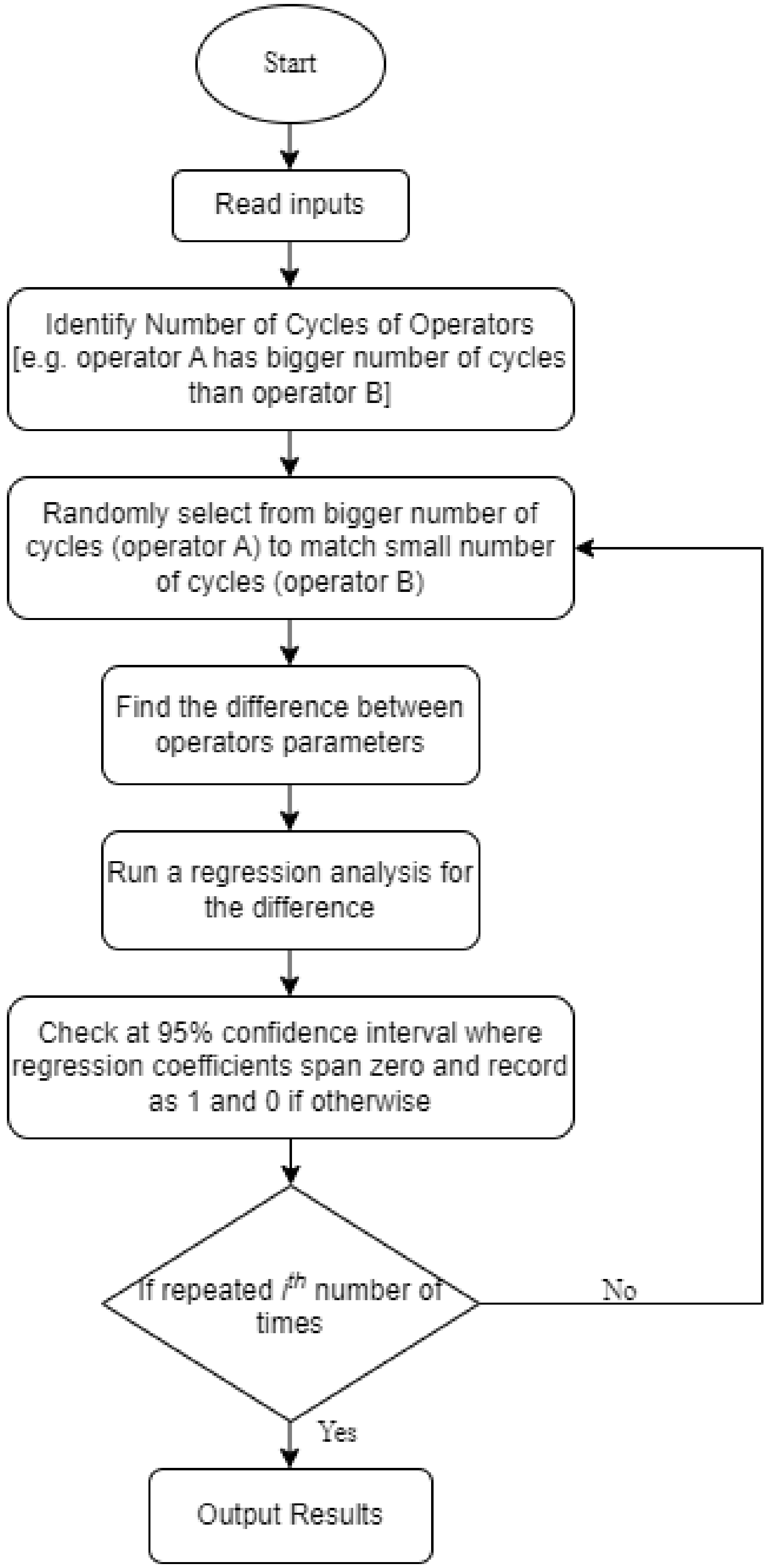

Figure 5.5 The base function algorithm developed to compare operators

The function requires three inputs: **OprA, Opr B, and** *i*. **OprA** is a matrix of *n*1 samples (i.e., number of rows indicating cycles) and *k* features (i.e., columns indicating correlated and dependent parameters). **OprB** is a matrix of *n*2 samples (i.e., number of rows indicating cycles) and *k* features (i.e., columns indicating correlated and dependent parameters). These two matrices are expected to have identical numbers of columns, signifying they have the same set of *k* features (correlated and dependent parameters).

However, the number of samples (cycles) in each dataset will most likely differ (Table 4.5). The third input, *i*, is a scalar indicating the number of iterations for the random sampling procedure. The function outputs a matrix **p** of size (k-1 × i). Each element of **p** is a binary number, with ‘1’ indicating that the confidence interval of the regression coefficient does not contain zero and ‘0’ indicating otherwise. In each iteration, the function checks which dataset has fewer rows (cycles) and sets it as the “short data.” It then randomly samples a subset of rows from the “long data” that matches the size of “short data” using the ‘datasample’ function of the Statistics and Machine Learning Toolbox in MATLAB and calculates the differences between the two datasets’ corresponding rows.

For example, when comparing operator *j* and operator *k* (Table 5.7), operator *j* has more cycles (40 cycles) than operator *k* (25 cycles). Random selection is made to have an equal number of cycles for both operators *j* and *k*. From operator *j*, 25 cycles are selected randomly to match the 25 cycles of operator *k*, and the difference between their corresponding parameters is estimated (Table 5.8). It is important to note that operators *j* and *k* have the same number of variables or parameters but different numbers of cycles.

Table 5.7 Operator *j* and operator *k* data pattern

	Operator j	Operator k
Cycle 1	X_{j1}	X_{k1}
Cycle 2	X_{j2}	X_{k2}
Cycle 3	X_{j3}	X_{k3}
	.	.
	.	.
Cycle 25	X_{j25}	X_{k25}
	.	
	.	
	.	
Cycle 40	X_{j40}	

Next, it performs a linear regression using the difference in the dependent variable as the response and the differences in the independent variables as the predictors. The function then checks if the 95% confidence intervals of the regression coefficients for the independent variables span zero. Finally, the base function repeats the process *i* times and stores the results in the **p** matrix. This process helps to test whether differences in the dependent variable are consistently associated with differences in the independent variables across multiple random subsets of the data.

Table 5.8 Difference between operators *j* and *k* parameters in matrix form

$$
\begin{array}{l}
\\
\text{Cycle 1}\\
\text{Cycle 2}\\
\text{Cycle 3}\\
\\
\\
\text{Cycle 25}
\end{array}
\begin{array}{c}
\begin{array}{cccccc}
\quad\Delta\eta\quad & & & \Delta var & &
\end{array}\\
\left[\begin{array}{c|ccccc}
\eta_{j1}-\eta_{k1} & var_{j11}-var_{k11} & var_{j12}-var_{k12} & . & . & var_{j1f}-var_{k1f}\\
\eta_{j2}-\eta_{k2} & var_{j21}-var_{k21} & var_{j22}-var_{k22} & . & . & var_{k2f}-var_{k2f}\\
\eta_{j3}-\eta_{k3} & var_{j31}-var_{k31} & var_{j32}-var_{k32} & . & . & var_{j3f}-var_{k3f}\\
. & . & . & . & . & .\\
. & . & . & . & . & .\\
\eta_{j25}-\eta_{k25} & var_{j251}-var_{k251} & var_{j252}-var_{k252} & . & . & var_{j25f}-var_{k25f}
\end{array}\right]
\end{array}
$$

The second (main) function takes multiple datasets and compares every pair of them using the base function. For example, ten pairs of operators would be required for comparison using the base function when five operators are under consideration (Equation 5.1). This function applies multiple linear regression analyses for each dataset pair and counts the number of times each independent variable is significant. This function takes a variable number of inputs, of which the first is i, the number of iterations for random sampling in the base function. The remaining input contains datasets for the operators under consideration represented as a matrix (at least two operators). Figure 5.6 presents the simplified flowchart of the main algorithm.

The function produces two outputs: **resultsMatrix** and **resultsMatrixNamed**. 'resultsMatrix' is a matrix where each row contains the count of significant independent variables for a specific comparison (for example, operator j and operator k). The size of this matrix is (numComparisons × numFeatures), where numComparisons is the total number of comparisons made (which is calculated by the formula in Equation 5.1), and numFeatures is the number of independent variables (9 in this case).

'resultsMatrixNamed' is a cell array containing dataset pair names in the first column and corresponding resultsMatrix entries in the remaining columns. Then it starts to perform comparisons for each pair of datasets. For each pair, it calls the base function, which returns a binary matrix indicating which independent variables are significant (represented by 1). It then sums up each row of this matrix to get the number of times each independent variable is significant. These counts are stored in the **resultsMatrix** and **resultsMatrixNamed.** The process repeats until all pairs of datasets have been compared.

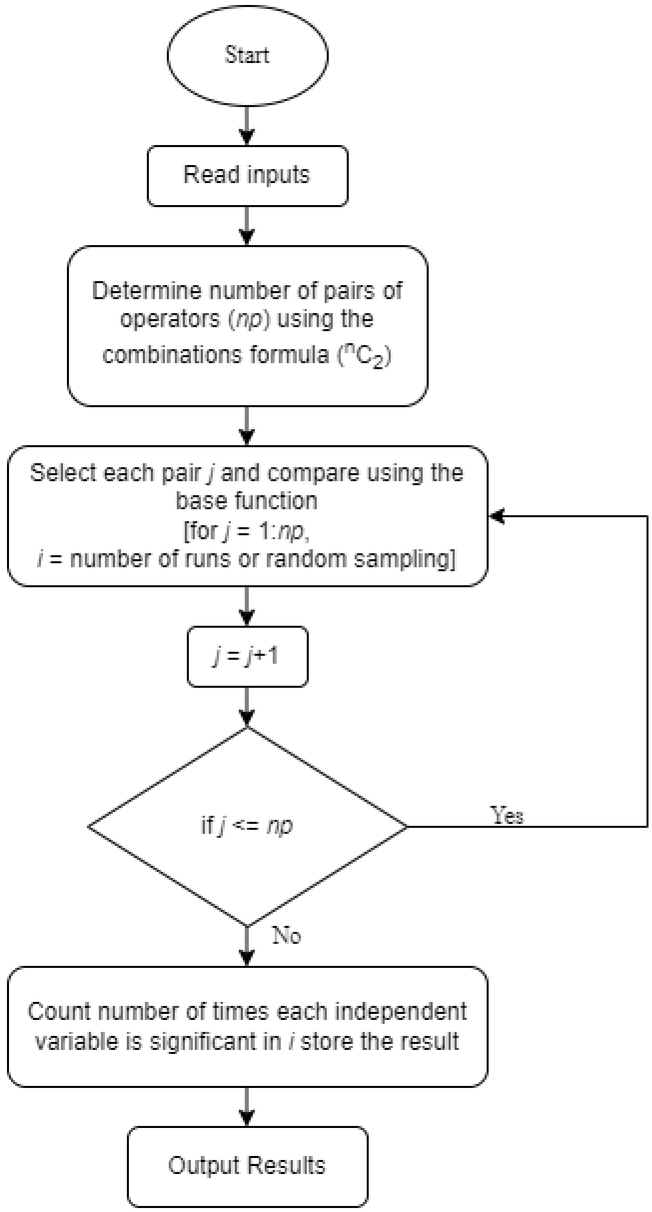

Figure 5.6 A simplified flowchart showing the main algorithm for the difference regression analysis

Notably, this researcher implements multiple iterations in the regression analysis to mitigate the impact of randomness associated with sampling. Conducting the regression process ‘i’ times enables the method to capture a broader range of possible sample combinations from the given datasets. This repetitive process ensures effective unbiased sampling that provides a more robust and reliable representation of the genuine underlying relationships in the data. This researcher bolsters the results' reliability and strengthens confidence in the derived conclusions by incorporating this iterative process.

In this study, nine of the extracted parameters demonstrate a correlation with energy efficiency. These include cycle time, swing-in time, digging time, swing-out time, dump height, payload, boom angle, stick angle, and swing-in angle. Furthermore, eight operators, labeled Operators A to H, were selected for the difference regression analysis. To ensure a robust and meaningful analysis, this researcher set the number of iterations to 30. The researcher conducted the difference regression analysis using the available data and adopting the difference regression analysis approach. The results of this thorough and rigorous analysis are presented in Table 5.9.

The study considers eight operators - Operator A through Operator H for the difference regression analysis. Given that the difference regression analysis is based on the pairwise comparison, the number of pairwise comparisons for the eight operators under study is 28 per Equation 5.1. An essential aspect of the analysis in this study involves assigning a binary number to the variables in the regression analysis based on a 95% confidence estimate of the importance of a variable in explaining the differences in energy per unit loading rate of the 30 iterations.

The criterion for this binary assignment involves the proportion of the total number of times a variable is significant (i.e., the regression coefficient is nonzero at 95% confidence level) within the 30 iterations. If a variable appears as significant in an amount greater than or equal to 0.95 times the total number of iterations (95% confidence of 30), the variable is assigned a binary value of 1. Otherwise, it receives a binary value of 0. This procedure, therefore, results in a binary representation of the significance of each variable across the iterations, providing an effective means to understand the impact of each variable on the model (Table 5.10).

Table 5.9 Regression analysis results for 28 operator comparisons (30 iterations each)

Pairs	Cycle time	Swing in time	Digging time	Swing out time	Dump height	Payload	Boom angle	Stick angle	Swing in angle
A,B	11	0	3	0	1	30	30	3	2
A,C	2	1	25	27	1	30	30	7	0
A,D	5	0	7	5	12	30	29	5	2
A,E	2	0	1	4	3	30	19	7	4
A,F	1	1	15	7	2	30	28	4	0
A,G	6	2	10	4	1	30	25	1	4
A,H	11	1	3	2	0	30	14	7	1
B,C	1	1	26	29	8	30	29	1	1
B,D	1	11	10	1	1	30	30	0	1
B,E	1	0	0	2	0	30	25	6	8
B,F	2	5	6	6	1	30	30	2	1
B,G	1	0	13	6	0	30	27	1	7
B,H	5	0	4	1	0	30	23	4	2
C,D	1	3	29	27	1	30	30	3	0
C,E	1	0	13	28	4	30	21	7	6
C,F	1	2	27	30	0	30	30	4	1
C,G	1	0	29	29	2	30	28	2	2
C,H	1	2	5	14	2	30	23	4	3
D,E	0	2	2	3	0	30	28	1	0
D,F	1	9	21	20	3	30	30	0	1
D,G	0	1	27	13	1	30	30	1	0
D,H	1	1	1	0	1	30	27	1	0
E,F	0	0	4	3	0	30	21	7	8
E,G	0	3	0	6	0	30	14	8	11
E,H	2	0	1	2	3	30	9	6	4
F,G	1	0	27	19	3	30	27	2	3
F,H	2	3	0	7	1	30	23	1	2
G,H	3	2	5	0	0	30	24	2	8

To illustrate this procedure, consider the variable cycle time as an example. If, in a comparison, ‘cycle time’ registers as significant in 10 out of the 30 iterations, ‘cycle time’ will be assigned a binary value of 0. This is because 10 is less than the 28.5 thresholds (95% confidence of 30 iterations) and the value of 1 if otherwise.

This researcher then proceeds to perform this binary assignment for each variable across all comparisons. The subsequent analysis then allows for a precise evaluation of which variables consistently impact the model with a 95% confidence interval. These identified variables subsequently offer the most promising avenues for further study and potential optimization.

Furthermore, the same methodology calculates each variable’s significance percentage. This involves summing up all the instances of ‘1’ after assigning the binary 0/1 across all comparisons (representing the times when a variable was significant meets the thresholds criterion). The researcher estimates the percentage of this sum with respect to the total number of comparisons (28 in this study). The percentage of the significance of a variable signifies the chance (probability) of the variable being responsible for the differences in operators’ energy efficiencies (Table 5.10). This approach allows for quantifying the influence of each variable in the model. It provides an adequate measure of the relative importance of each variable in the context of this analysis.

The results of the binary assignment and subsequent analysis reveal valuable insights into the influence of various operational parameters on energy efficiency. Payload is the most influential factor, with a 100% chance of explaining differences in energy per unit loading rates, or energy efficiencies, across all operator comparisons. This suggests that payload significantly impacts the energy efficiency of the operation.

Assuming all other conditions remain constant, operators with higher payloads will likely have higher energy efficiencies (i.e., lower energy per unit loading rates).

Table 5.10 Probability result after assigning binary numbers 0/1

	Cycle time	Swing in time	Digging time	Swing out time	Dump height	Payload	Boom angle	Stick angle	Swing in angle
A,B	0	0	0	0	0	1	1	0	0
A,C	0	0	0	0	0	1	1	0	0
A,D	0	0	0	0	0	1	1	0	0
A,E	0	0	0	0	0	1	0	0	0
A,F	0	0	0	0	0	1	0	0	0
A,G	0	0	0	0	0	1	0	0	0
A,H	0	0	0	0	0	1	0	0	0
B,C	0	0	0	1	0	1	1	0	0
B,D	0	0	0	0	0	1	1	0	0
B,E	0	0	0	0	0	1	0	0	0
B,F	0	0	0	0	0	1	1	0	0
B,G	0	0	0	0	0	1	0	0	0
B,H	0	0	0	0	0	1	0	0	0
C,D	0	0	1	0	0	1	1	0	0
C,E	0	0	0	0	0	1	0	0	0
C,F	0	0	0	1	0	1	1	0	0
C,G	0	0	1	1	0	1	0	0	0
C,H	0	0	0	0	0	1	0	0	0
D,E	0	0	0	0	0	1	0	0	0
D,F	0	0	0	0	0	1	1	0	0
D,G	0	0	0	0	0	1	1	0	0
D,H	0	0	0	0	0	1	0	0	0
E,F	0	0	0	0	0	1	0	0	0
E,G	0	0	0	0	0	1	0	0	0
E,H	0	0	0	0	0	1	0	0	0
F,G	0	0	0	0	0	1	0	0	0
F,H	0	0	0	0	0	1	0	0	0
G,H	0	0	0	0	0	1	0	0	0
Total	0	0	2	3	0	28	10	0	0
%	0	0	7	11	0	100	36	0	0

Next, the boom angle appears to explain differences in energy efficiencies 36% of the time. This is considerably lower than the payload's influence but still notable. It indicates that the overall displacement of the boom during a cycle differs among the operators and is a significant contributor (at least explains differences between 36% of operator pairs) to differences in operator energy efficiency. However, as it is less consistent across comparisons, the boom angle might be subject to other influencing factors, such as the operator's technique or the specific task requirements.

Digging time and swing-out time show some potential for causing differences in energy efficiencies, with influences of 7% and 11%, respectively. However, their impacts are considerably lower than those of the payload and the boom angle, suggesting that while these time factors can contribute to differences in energy efficiencies, their effects are less consistent across different operator comparisons.

The remaining parameters, including cycle time, swing-in time, dump height, stick angle, and swing-in angle, show no explanatory power in explaining differences in energy efficiencies across the considered comparisons. This indicates that while these factors might be crucial for other aspects of the operation, they do not appear to contribute significantly to differences in energy efficiencies across operators. It is important to note that the lack of significance for certain variables does not necessarily mean they have no impact on energy efficiencies. The absence of consistent significance across pairwise comparisons just means these variables are not the source of differences between operator energy efficiencies (as defined with the data in this research).

The findings from this research underscore the critical notion that the specific variables identified as significant depend on the chosen measure of operator performance.

In this instance, the measure utilized is energy per unit loading rate, calculated as (total energy × cycle time)/payload (Equation 2.2). The total energy component in this formula comprises the energies expended by various parts of the hydraulic shovel, including the boom, stick, swing, and bucket. However, due to the limited nature of the available dataset, only the energy expended by the boom could be included in the total energy calculation (Equation 4.1). This limitation has potential implications for the identification of significant variables. Certain variables that may otherwise have been significant if the energies of the stick, swing, and bucket were included might not have been detected as such due to the current measure's specific focus on the boom energy. In other words, the dataset limitation might inadvertently mask the significance of some variables, causing them to be overlooked in this analysis.

Despite these data limitations, it is important to recognize the value of the approach adopted in this research. The developed algorithms, which can conduct complex statistical analyses, have proven effective and robust in working with the available data. Further, the researcher believes these algorithms possess the flexibility and adaptability to handle more comprehensive data, capturing all of the energies of a hydraulic shovel, if such data were to be made available. Therefore, while the current dataset's limitation is an acknowledged constraint of this research, it does not detract from the effectiveness of the overall approach. The inherent adaptability of the algorithms developed in this study ensures their readiness to provide even more nuanced insights when better and more comprehensive data becomes available. As such, this study provides valuable insights into the current dataset and paves the way for future investigations with improved data.

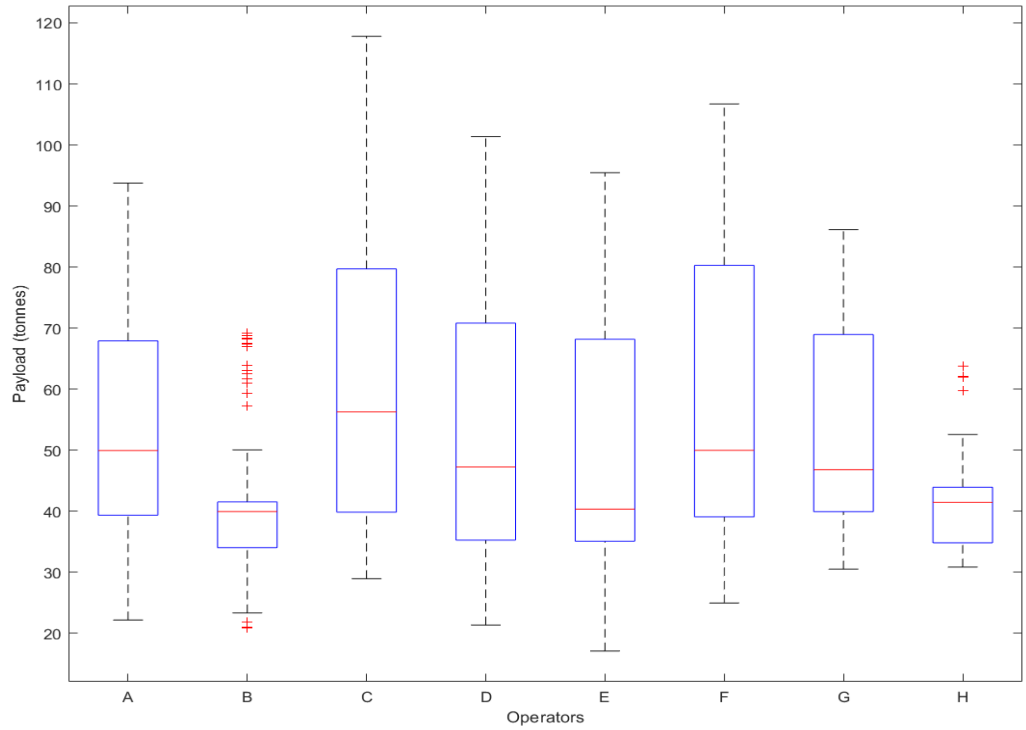

Figure 5.7 Plots showing the boxplots of payloads of operators[7]

An important finding from the analysis is that even with the limitations posed by the dataset, 'payload' variable consistently emerges as a distinguishing factor among operators in all the comparisons. This consistency points to 'payload' as a distinguishable factor in the performance of different operators. From a skeptical perspective, one might argue that since the payload is directly included in the computation of energy per unit loading rate, its impact is somewhat expected. As a counter illustration, consider cycle time. Despite being a parameter in the formula for calculating energy per unit loading

[7] The payloads for operators B and H are unusually skewed and the researcher's data review did not determine an explanation for this high skewness.

rate, the algorithm did not consistently identify cycle time as a significant variable that explains differences in energy per unit loading rate across all operators. This demonstrates that the significance of a variable is not merely determined by its presence in the performance measure but its variability among operators. Therefore, the payload's identification as a key factor indicates true differences among operators, not just a result of its role in calculating energy per unit loading rate. This underscores the strength and nuance of the algorithm developed by the researcher. Figure 5.7 shows the distribution of payloads for the different operators. The figure shows that there is clearly a wide variation in the distribution of payloads for each operator.

The benefit of the approach suggested in this thesis is that it provides "actionable" data that mine managers and engineers can use to make changes to improve energy efficiency. For example, the result showing that payload is important can be used to guide operators to improve overall energy efficiency and cost improvements. Here, let us use two operators from this study (Table 4.5), the operator with the best performance (lowest energy per unit loading rate), Operator G, and the one with the poorest performance (highest energy per unit loading rate), Operator E, to estimate and illustrate the possible energy efficiency and cost savings if management can get low performing operators to improve their payloads per cycle. APPENDIX B-1 estimates the potential energy efficiency and cost savings. Using Operator G as the standard to improve Operator E, the initial production rate of Operator E is about 81% of that of Operator G. By increasing the payload of Operator E from 47.18 tonnes to the level of Operator G (53.15 tonnes), the loading rate of Operator E increases to approximately 92% of that of Operator G, marking an increase of around 11%. Simultaneously, Operator E's energy per unit loading rate decreases from 10,610 kJsecs/tonne to 9418.73 kJsecs/tonne after the payload improvement, implying an energy efficiency enhancement of about 11%. Additionally, the cost of energy per unit loading rate for Operator E decreases from $0.28 to $0.24 after the payload improvement, which is a decline of about 14% in the cost of energy per unit loading rate. This illustrates how mine managers can use the data from the approach suggested in this thesis to optimize operator practices. In this case, payload optimization can improve energy and cost savings.

In the conducted analysis, digging time and swing out time did show up in the comparison of operators, but not as consistently as payload. This intermittent presence suggests that these factors may have some degree of influence on the energy per unit loading rate. However, they did not reach a level of statistical significance to draw conclusive decisions about their influence on energy efficiency. This partial visibility of these variables could result from the limited available dataset. With a more comprehensive dataset, it might be possible to better understand the role these variables play in differentiating operator performance. The inconsistencies found in the current analysis may well be artifacts of data limitations rather than reflections of the actual operational dynamics. In light of this, a more exhaustive dataset could provide a more nuanced picture, potentially revealing other key performance-differentiating variables and thereby substantiating or amending the preliminary observations made about the impact of digging time and swing-out time.

The researcher has made a significant contribution by successfully developing algorithms to apply an approach introduced by Abdi-Oskouei (2013) to the context of hydraulic shovels and this specific sensor dataset. Abdi-Oskouei’s original methodology, applied to draglines, provided a foundation the researcher has adapted and expanded upon. Creating these algorithms required a nuanced understanding of the original approach, the specifics of hydraulic shovels' operation, and the nature of the sensor data collected from a different monitoring system than what Abdi-Oskouei used in her research. This researcher has taken the existing knowledge from the academic domain and applied it effectively to new, complex, and practical situations, effectively bridging the gap between theory and practice.

Moreover, the successful development and implementation of these algorithms reveal that they can robustly handle this type of sensor data, offering a new toolset to analyze and understand the effect of operator practices (precisely cycle time and payloads) on energy efficiency of hydraulic shovels. By harnessing the statistical power of regression analysis, these algorithms enable nuanced and detailed insights to be drawn about factors influencing energy efficiency of hydraulic shovels, opening avenues for operator performance improvement.

It is worth noting that while the results are grounded in rigorous statistical analysis, the methodology's true strength comes from its adaptability. This ability to apply a refined version of Abdi-Oskouei's approach to different contexts underscores its versatility and the potential for further applications in related fields. The researcher's successful adaptation of this approach promises to add value to the study of hydraulic shovels and broader discussions on energy efficiency in mining and machine operation optimization.

5.3. SUMMARY

The primary objective of this section was to identify the key parameters responsible for the variations in energy efficiencies of operators. The analysis began by comparing the energy efficiencies of operators using the measure of energy per unit loading rate. The study reveals significant variations in energy efficiencies among the operators through rigorous analysis and hypothesis testing. The findings confirm that the energy efficiencies of operators are not uniform and highlight the need for further investigation into the underlying factors contributing to these variations.

Additionally, this researcher conducts a correlation analysis to examine the relationships between energy efficiencies and various operational parameters. The results reveal significant associations between energy efficiencies and cycle time, digging time, swing-out time, dump height, boom angle, and stick angle positively correlate with energy efficiencies. These findings suggest that these parameters may play a role in differentiating operator performance regarding energy efficiency. Payload shows a strong negative correlation with energy efficiencies (rho = -0.6472, p-value = 0.0000), indicating that payload significantly affects energy efficiencies. Other parameters, such as swing-in time, bucket angle, swing-in angle, and swing-out angle, exhibit weaker correlations that are not statistically significant. These correlation results provide valuable insights into the potential influence of specific operational parameters on energy efficiencies and can serve as a basis for further investigation.

In order to identify the key parameters responsible for the differences in energy efficiency among operators, this researcher employed a difference regression analysis to determine the significance of the correlated parameters in explaining differences in energy efficiencies of operators. The developed algorithms reveal that payload consistently emerged as a significant variable (100% of the time) in distinguishing operator energy efficiencies. The researcher acknowledges that although the energy per unit loading rate is a function of payload, the variations in payload among operators further validate its importance.

Moreover, the limited dataset used in this study may have constrained the identification of other influential parameters. The approach and algorithms developed in this research are highly adaptable. With better data capturing all the energies of hydraulic shovels, these algorithms would provide a more comprehensive understanding of the factors explaining differences in energy efficiencies among operators.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1. OVERVIEW

Hydraulic shovels are essential and dominant equipment in mining because of their efficiency and productivity. However, a crucial aspect often overlooked is the impact of operators' practices on energy consumption and productivity. With the mining sector accounting for substantial energy usage and resultant carbon emissions, enhancing energy efficiency in these domains is urgent. Despite the recognition that operator practices influence the efficiency of loading machines in previous studies, research has sparingly quantified this impact, especially for hydraulic shovels. This work's novelty lies in addressing this gap by employing telemetry technology and advanced data analysis to investigate how varying operator practices contribute to the differences in energy consumption and productivity of hydraulic shovels.

The research primarily aims to achieve two objectives: (1) develop algorithms that can extract significant data from shovel telemetry for comprehensive statistical analysis, and (2) test the hypothesis that operator practices and skills have a notable effect on hydraulic shovel energy efficiency. The researcher developed several algorithms in this study to achieve the first objective. The initial algorithm sampled cycles and identified sub-cycles from telemetry data collected from a 40.5 yd^3 bucket hydraulic shovel. The researcher then developed subsequent algorithms to extract cycle-based data for significant parameters used in the analysis.

Regarding the second objective, the researcher first tested the hypothesis to statistically ascertain whether the difference in average energy efficiencies of operators is significant. Subsequently, this study conducted a correlation analysis to identify which parameters correlate with energy efficiency. The final stage of the analysis involved developing a difference linear regression model that regressed the difference in correlated parameters between operators against the difference in their energy efficiencies.

6.2. CONCLUSIONS

The following conclusions have been drawn from this work:

- This research has successfully developed an algorithm that can sample “ideal” cycles (i.e., cycles with clear demarcation of stages, no bench clean up, and where the swing is in the direction where the operator’s vision is not unobstructed by the boom) from the monitoring file from the commercial hydraulic shovel monitoring system used in this research. Manual validation using visual inspections of plots of swing, boom, and stick angular displacements and the commercial partner’s state enum variable to validate the cycle sampling algorithm show that the developed algorithm is 98% accurate.
- Based on the nature of the monitoring data and a review of the literature focused on finding factors that are likely to explain differences in operator energy per unit loading rate, this work selected 12 KPIs for each cycle: cycle time and cycle time components (swing-in, digging, and swing-out), payload, energy use, dump height and boom, bucket, swing-in, swing-out, and stick angles. Additionally, this work developed an algorithm to extract operator identities to facilitate separating data by operator. This work has successfully

developed algorithms, implemented in MATLAB as MATLAB functions, to extract these KPIs from the hydraulic shovel monitoring data.

- The study proposed an approach to selecting which operators should be included in the study. This study used the mean standard error statistic of the energy efficiencies of operators as the criterion with a cut-off value of 32 to decide which operators to include in the analysis. Based on this analysis, the researcher selected eight (8) operators to include in the study.
- The study revealed significant variations in energy efficiencies among the operators' p-values for the two tests (Welch's ANOVA and Kruskal-Wallis tests) effectively zero. The findings confirm that the energy efficiencies of operators are not the same and highlight the need for further investigation into the underlying factors contributing to these variations.
- The results of correlation analysis to examine the relationships between energy efficiencies and various operational parameters revealed significant associations between energy efficiency and cycle time, digging time, swing-out time, dump height, boom angle, stick angle, and payload. Cycle time, digging time, swing-out time, dump height, boom angle, and stick angle positively correlate with energy efficiency. Conversely, payload showed a strong negative correlation with energy efficiency (rho = -0.6472, p-value = 0.0000), indicating that payload significantly affects energy efficiency. These findings suggested that these parameters may play a role in differentiating operator performance regarding energy efficiency. Other parameters, such as

swing-in time, bucket angle, swing-in angle, and swing-out angle, exhibit weaker correlations that are not statistically significant.

- This study employed a difference linear regression analysis model to develop an algorithm that determines the significance and quantifies the correlated parameters' impact on differences in energy efficiency. Payload consistently emerged as a significant variable (100% of the time) in distinguishing operator energy efficiencies, while variables such as boom angle, swing-out time, and digging time were less consistent in explaining differences in operator energy efficiencies.
- The limited dataset used in this study may have constrained the identification of other influential parameters. The developed algorithms are adaptable and can be applied with better data to capture the total energy of hydraulic shovels, which would provide a more comprehensive understanding of the factors influencing energy efficiencies among operators.

6.3. CONTRIBUTIONS OF THE WORK

The work has made the following contributions to the body of knowledge, science, and mining engineering practice:

- This study is a pioneering effort to discern the impact of operator practices on the energy efficiency of hydraulic shovels. It marks the very first initiative toward quantifying this influence with field data. Before this investigation, empirical study on the role of operator practices in driving energy efficiency in this field remained largely unexplored and undefined. This research is a pioneering contribution to the current body of knowledge, throwing light on the nuances of how operator behavior can significantly alter the energy efficiency of hydraulic shovels. This opens up a new dimension in the ongoing quest for sustainable and efficient energy usage in this sector.

- Another significant contribution of this study is developing a novel algorithm to extract valuable insights from telemetry data. This algorithm transcends conventional uses of this technology by offering a reliable, data-driven approach to understanding the intricate relationship between operator practices and the energy efficiency of hydraulic shovels. This algorithm hinges on the algorithm to sample cycles from the raw sensor files, which allows the other algorithms to extract cycle-based information for further analyses.
- This study provides crucial data that can serve as a valuable resource for operator improvement. It enables an operator in the field to be educated and trained about their current production efficacy, instigating conscious improvements. The insights derived from this study provide an intelligent guide for operators, facilitating a tangible performance comparison. While mines typically employ monitoring systems to track payload, this study introduces a novel approach to leverage this data to initiate a constructive dialogue with operators. It allows for meaningful comparison of operator performances, helping them understand any deficiencies in their current production rates and their impact on energy efficiency. As illustrated before and shown in Appendix B-1, it is possible to decrease energy costs by 14%

while achieving a similar loading rate using the data generated by this approach to guide operators.

- An additional benefit of this study's approach is that it enhances energy efficiency without incurring any additional expenditure. The focus here is on improving operator practices, particularly regarding payload optimization, using the existing telemetry data. This cost-effective method of boosting energy efficiency ensures improved production rates and energy savings, underlining the concept that significant improvements can be made without the need for substantial monetary investment. It is a strategy of 'doing more with less,' focusing on improving practices with the resources at hand.

6.4. RECOMMENDATIONS

❖ The author recommends this for mining engineering practice:

- Recognizing the vital role that the shovel payload plays in differentiating various operators, it is crucial to highlight the need for adequate operator training in maximizing bucket fill factor to improve payloads. A well-trained operator will be adept at optimizing the bucket fill, which, in turn, will increase the payload capacity and directly enhance the production rate. This increase in production rate boosts output and improves energy efficiency. A larger payload implies that more material can be loaded or produced for the same amount of energy, lowering the energy per unit loading rate. This is a significant advantage considering the escalating global concerns around energy consumption and the necessity for sustainable usage. Moreover, there is a direct economic benefit attached to this. A decrease in energy per unit loading rate translates to a reduction in energy cost per unit of production, substantially lowering the overall operational costs. Thus, it is of utmost importance for organizations to invest in effective operator training guided by field data (so there is no ambiguity in what is necessary). It results in increased payloads and production rates, boosts energy efficiency, and reduces operational costs, presenting a win-win situation for the operator and the organization.

❖ This author also recommends the following for future work:

- Future work should consider expanding the dataset to capture more comprehensive data for each operator. The current data is limited and might not fully reveal the significant influence of variables such as boom angle, digging, and swing-out times, and a broader dataset can help uncover these nuanced differences between operators.
- Future research should incorporate additional measures of energy use. The current calculation of energy per unit loading rate only factors in the energy used by the boom. Including energy expended by the stick, swing, bucket, and other components can provide a more comprehensive understanding of energy efficiency.
- Although this study used a linear regression model to analyze mixed data types (both linear and circular), there is potential for even more accurate insights if future research incorporates a linear-circular regression model. This approach, as suggested by Fisher et al. (1992), Kim et al. (2018),

Kato et al. (2008), and Mohammed et al. (2021) could be more reliable in their inferences. Therefore, future studies should endeavor to apply these linear-circular regression models for potentially enhanced results and deeper insights into hydraulic shovel operator efficiency.

- Given the apparent influence of payload on operator differences, further research should investigate the reasons behind varying payloads between operators. The insights gleaned could contribute significantly to operator training and performance optimization strategies. Any effort that helps operators fill the bucket better (and thus improve payload) will improve their energy efficiency.
- Future research should consider delving deeper into operator behavior and decision-making processes under challenging digging situations. This study has touched upon the operator practices influencing efficiency and productivity, but there is room to understand how operators navigate complex digging situations. Investigating this aspect could yield invaluable insights into operator skills, learning mechanisms, and adaptability, potentially leading to further optimization of hydraulic shovel operation and a significant increase in energy efficiency.

APPENDIX A.

GRAPHICAL PLOTS FOR NORMALITY TEST

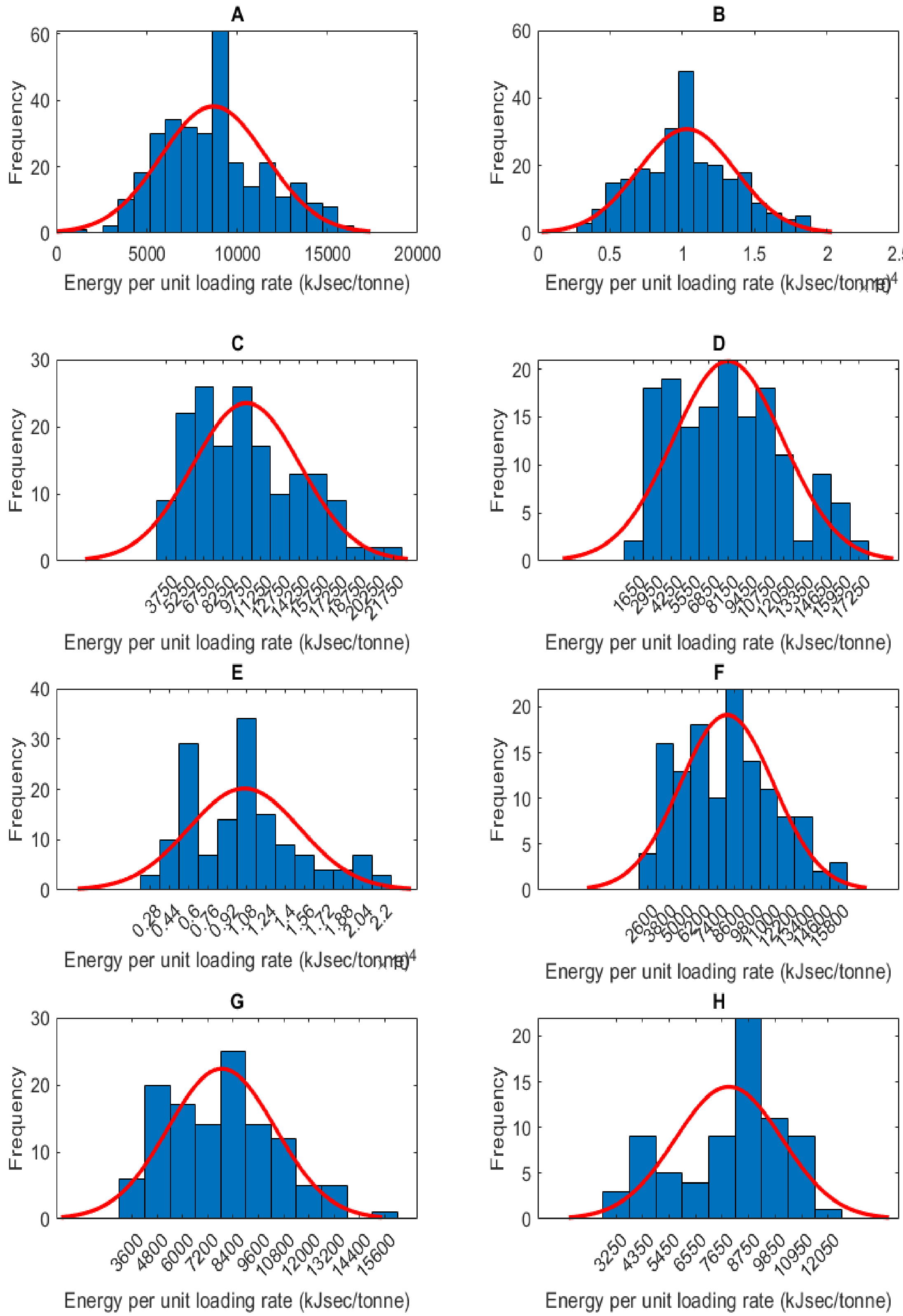

Figure A.1. Distributions of energy per unit loading rates of operators

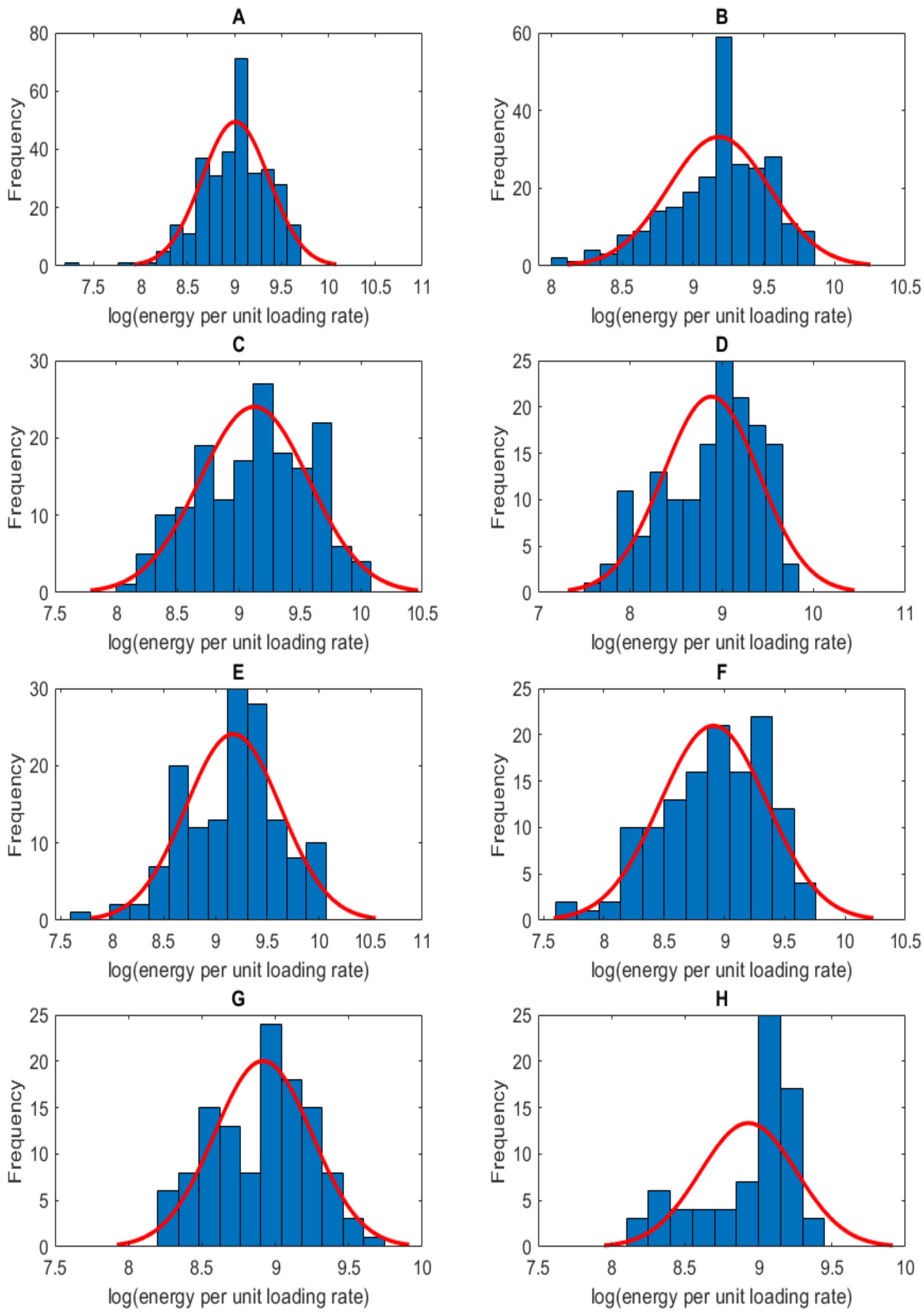

Figure A.2. Distributions of log-energy per unit loading rates of operators

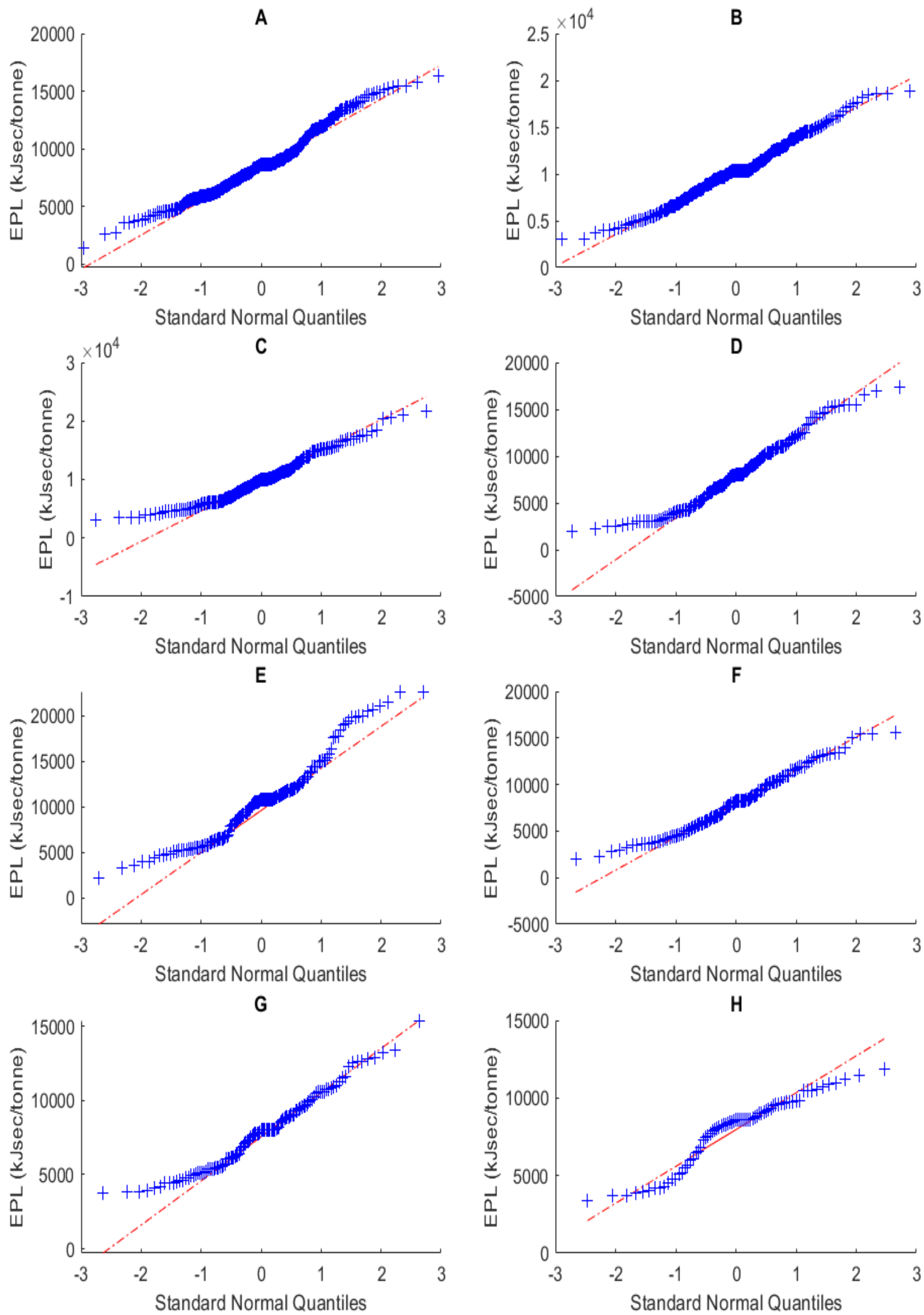

Figure A.3. Q-Q plots of energy per unit loading rates of operators

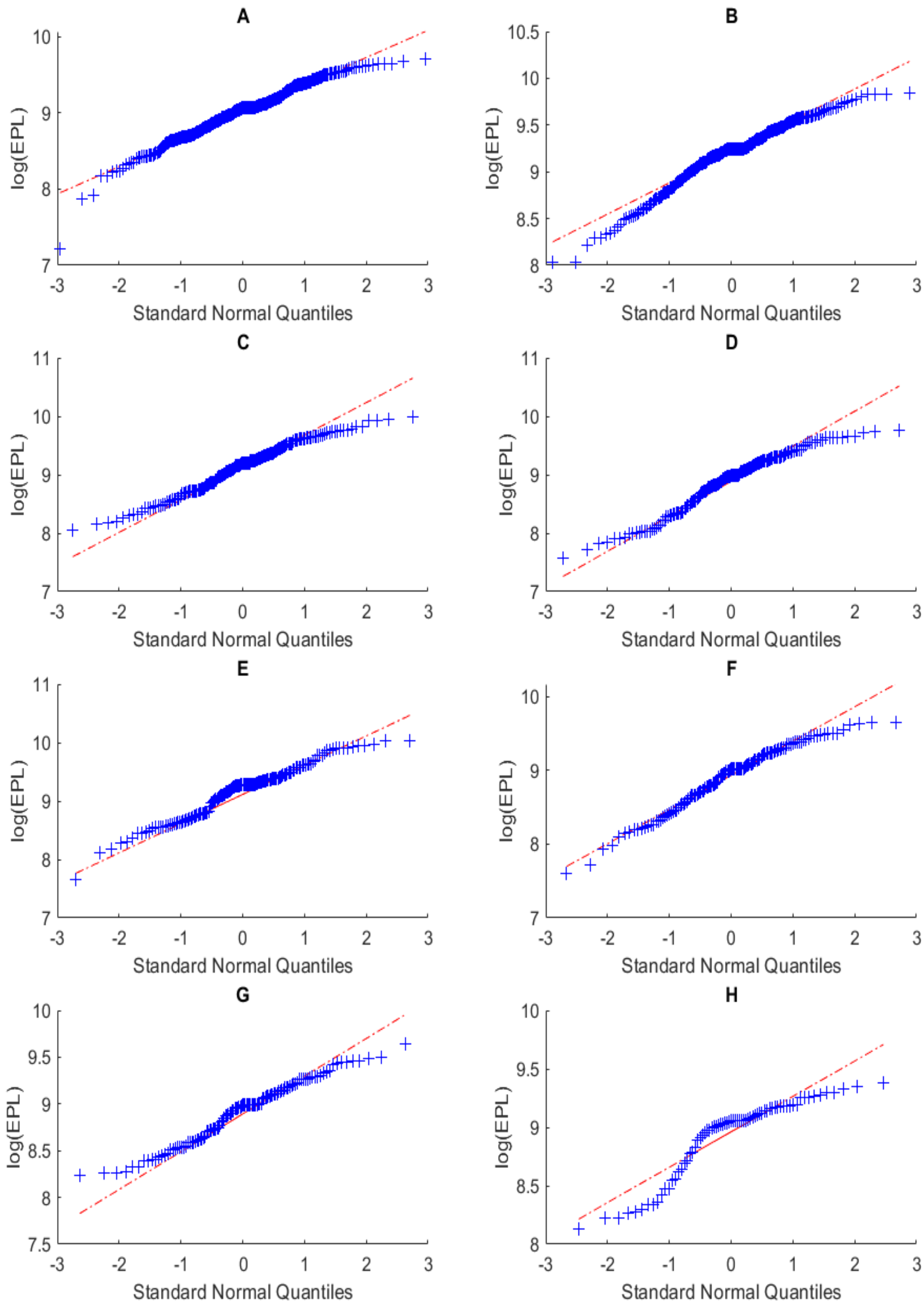

Figure A.4. Q-Q plots of log-energy per unit loading rates of operators

APPENDIX B.

POTENTIAL ENERGY SAVINGS ESTIMATES

ENERGY EFFICIENCY IMPROVEMENT ESTIMATES

APPENDIX B shows the analysis done to estimate the possible energy savings if the mine was to get operator E (worst performance) to produce (load) at the level of operator G (highest performance). The energy price of "$3.806/gal ($0.000025976/kJ)" used in the analysis is based on the diesel fuel price obtained from the U.S. Energy Information Administration (U.S. Energy Information Administration, 2023).

Table B-1. Potential energy efficiency and energy cost savings estimates

	Best	Worst	
Parameters	Operator G	Operator E	E's payload Improved to G's Payload
Energy per cycle (kJ)	15,102	16,575	16,575
Cycle time (sec)	27.65	30.20	30.20
Payload per cycle (tonnes)	53.15	47.18	53.15
Production rate (tonnes/sec)	1.92	1.56	1.76
Energy per unit loading rate	7,855	10,611	9,419
% of best operator's loading rate	100%	81.25%	91.53%
Energy price ($/kJ)	2.6×10^{-5}	2.6×10^{-5}	2.6×10^{-5}
Energy cost for loading rate	$0.20	$0.28	$0.24
Energy costs (savings)		14.29%	

BIBLIOGRAPHY

Abdi-Oskouei, M. (2013). *Methods for evaluating effect of operators on dragline energy efficiency* [Missouri University of Science and Technology]. https://scholarsmine.mst.edu/masters_theseshttps://scholarsmine.mst.edu/masters_theses/7141

Abdi-Oskouei, M., & Awuah-Offei, K. (2014). Statistical methods for evaluating the effect of operators on energy efficiency of mining machines. *Transactions of the Institutions of Mining and Metallurgy, Section A: Mining Technology*, *123*(4), 175–182. https://doi.org/10.1179/1743286314Y.0000000067

Abdi-Oskouei, M., & Awuah-Offei, K. (2016). A method for data-driven evaluation of operator impact on energy efficiency of digging machines. *Energy Efficiency*, *9*(1), 129–140. https://doi.org/10.1007/s12053-015-9353-3

Acaroglu, O., Ozdemir, L., & Asbury, B. (2008). A fuzzy logic model to predict specific energy requirement for TBM performance prediction. *Tunnelling and Underground Space Technology*, *23*(5), 600–608. https://doi.org/10.1016/J.TUST.2007.11.003

Aguirre-Jofré, H., Eyre, M., Valerio, S., & Vogt, D. (2021). Low-cost internet of things (IoT) for monitoring and optimising mining small-scale trucks and surface mining shovels. *Automation in Construction*, *131*, 103918. https://doi.org/https://doi.org/10.1016/j.autcon.2021.103918

Akkaş, M. A. (2018). Using wireless underground sensor networks for mine and miner safety. *Wireless Networks*, *24*(1), 17–26. https://doi.org/10.1007/s11276-016-1313-0

Andreev, R., Joseph, T., Sammut, J., & Curley, M. (2017). *HYDRAULIC EXCAVATOR VS. ROPE SHOVEL PERFORMANCE*.

Awuah-Offei, K. (2016). Energy efficiency in mining: A review with emphasis on the role of operators in loading and hauling operations. *Journal of Cleaner Production*, *117*, 89–97. https://doi.org/10.1016/j.jclepro.2016.01.035

Awuah-Offei, K. (2018a). Energy Efficiency in Cable Shovel Operations. In *Green Energy and Technology* (pp. 147–157). https://doi.org/10.1007/978-3-319-54199-0_8

Awuah-Offei, K. (2018b). *Green Energy and Technology Energy Efficiency in the Minerals Industry - Best Practices and Research Directions* (K. Awuah-Offei Editor, Ed.). Springer International Publishing. http://www.springer.com/series/8059

Awuah-Offei, K., & Frimpong, S. (2007). Cable shovel digging optimization for energy efficiency. *Mechanism and Machine Theory*, *42*(8), 995–1006. https://doi.org/10.1016/j.mechmachtheory.2006.07.008

Awuah-Offei, K., & Frimpong, S. (2011). Efficient Cable Shovel Excavation in Surface Mines. *Geotechnical and Geological Engineering*, *29*(1), 19–26. https://doi.org/10.1007/s10706-010-9366-9

Awuah-Offei, K., Frimpong, S., & Askari-Nasab, H. (2009). Formation excavation resistance modelling for shovel dippers. In *Int. J. Mining and Mineral Engineering* (Vol. 1, Issue 2).

Awuah-Offei, K., Osei, B., & Askari-Nasab, H. (2011). *Awuah-Offei K. et al. Modeling Truck-Shovel Energy Efficiency under Uncertainty*.

Babaei, M., & Hall, R. (2016). A Study of Digging Productivity of an Electric Rope Shovel for Different Operators. *Minerals*, *6*, 48. https://doi.org/10.3390/min6020048

Bakshi, U. A., & Bakshi, A. V. (2020). *Electronic Measurements and Instrumentation*. Technical Publications. https://books.google.com/books?id=DqkaEAAAQBAJ

BenSaida, A. (2009). *Shapiro-Wilk and Shapiro-Francia normality tests* (1.1.0.0 (R2023a)). MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/13964-shapiro-wilk-and-shapiro-francia-normality-tests

Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. *Journal of Statistical Software*, *31*. https://doi.org/10.18637/jss.v031.i10

Berens, P. (2023). *Toolbox for Circular Statistics (Directional Statistics)* (1.21.0.0). MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/10676-circular-statistics-toolbox-directional-statistics

Best, D. J., & Roberts, D. E. (1975). Algorithm AS 89: The Upper Tail Probabilities of Spearman’s Rho. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *24*(3), 377–379. https://doi.org/10.2307/2347111

Bhattacharya, S., & SenGupta, A. (2009). Bayesian Analysis of Semiparametric Linear-Circular Models. *Journal of Agricultural, Biological, and Environmental Statistics*, 36–65.

Biau, D. J. (2011). In Brief: Standard Deviation and Standard Error. *Clinical Orthopaedics and Related Research®*, *469*(9), 2661–2664. https://doi.org/10.1007/s11999-011-1908-9

Bouchard, J., Sbarbaro, D., & Desbiens, A. (2018). Plant Automation for Energy-Efficient Mineral Processing. In *Green Energy and Technology* (pp. 233–250). https://doi.org/10.1007/978-3-319-54199-0_13

Carvalho, M., Romero, A., Shields, G., & Millar, D. L. (2014). Optimal synthesis of energy supply systems for remote open pit mines. *Applied Thermal Engineering*, *64*(1), 315–330. https://doi.org/https://doi.org/10.1016/j.applthermaleng.2013.12.040

Caterpillar. (2020). *CAT ® HYDRAULIC MINING SHOVELS PAYLOAD MANAGEMENT 1 CAT ® HYDRAULIC MINING SHOVELS*. https://caterpillar.scene7.com/is/content/Caterpillar/CM20200828-1fe9e-68367

Caterpillar. (2023). *Hydraulic Mining Shovels/Excavators*. Caterpillar Inc. https://www.cat.com/en_US/products/new/equipment/hydraulic-mining-shovels.html

Chen, X., Wu, W., Wang, H., & Zhou, Y. (2010). Distributed monitoring and diagnosis system for hydraulic system of construction machinery. *Frontiers of Mechanical Engineering in China*, *5*(1), 106–110. https://doi.org/10.1007/s11465-009-0089-7

Chetty, B. A. (1982). TELEMETRY IN THE MINING INDUSTRY. *IETE Journal of Research*, *29*(8), 389–391. https://doi.org/10.1080/03772063.1983.11453100

Curran-Everett, D. (2017). Explorations in statistics: the assumption of normality. *Advances in Physiology Education*, *41*(3), 449–453. https://doi.org/10.1152/advan.00064.2017

Demirel, N. (2018). Energy-Efficient Mine Ventilation Practices. In K. Awuah-Offei (Ed.), *Energy Efficiency in the Minerals Industry: Best Practices and Research Directions* (pp. 287–299). Springer International Publishing. https://doi.org/10.1007/978-3-319-54199-0_16

Dindarloo, S. R., Siami-Irdemoosa, E., & Frimpong, S. A. (2016). Measuring the effectiveness of mining shovels. *Mining Engineering*, *68*, 45–50. https://www.researchgate.net/publication/292965108_Measuring_the_effectiveness_of_mining_shovels

Dong, L., Mingyue, R., & Guoying, M. (2017). Application of Internet of Things Technology on Predictive Maintenance System of Coal Equipment. *Procedia Engineering*, *174*, 885–889. https://doi.org/10.1016/J.PROENG.2017.01.237

Drezner, Z., Turel, O., & Zerom, D. (2010). A Modified Kolmogorov–Smirnov Test for Normality. *Communications in Statistics - Simulation and Computation*, *39*, 693–704.

Duarte, J., Rodrigues, F., & Branco, J. C. (2022). Sensing Technology Applications in the Mining Industry—A Systematic Review. *International Journal of Environmental Research and Public Health*, *19*(4). https://doi.org/10.3390/ijerph19042334

Esmaieeli-Sikaroudi, A. (2017). *Regressing over Linear-Circular Data Using a Mixture of Linear-Linear Regression Models*. Florida State University.

Fantini, A., Fiorucci, M., Martino, S., & Paciello, A. (2017). Investigating Rock Mass Failure Precursors Using a Multi-sensor Monitoring System: Preliminary Results From a Test-Site (Acuto, Italy). *Procedia Engineering*, *191*, 188–195. https://doi.org/10.1016/J.PROENG.2017.05.171

Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511564345

Fisher, N. I., & Lee, A. J. (1981). Nonparametric Measures of Angular-Linear Association. *Biometrika*, *68*(3), 629–636. https://doi.org/10.2307/2335445

Fisher, N. I., & Lee, A. J. (1992). Regression Models for an Angular Response. *Biometrics*, *48*, 665–677.

Gibbons, J. D. (1985). *Nonparametric statistical methods* (2nd ed., Vol. 65). https://openlibrary.org/books/OL2532185M/Nonparametric_statistical_inference

Guzman, M., Valenzuela, M., & Member, S. (2015). Integrated mechanical-electrical modeling of an ac electric mining shovel and evaluation of power requirements during a truck loading. *Cycle*, *51*, 2590–2599.

Hadjigeorgiou, J., & Poulin, R. (1998). Assessment of ease of excavation of surface mines. *Journal of Terramechanics*, *35*(3), 137–153. https://doi.org/10.1016/S0022-4898(98)00018-4

Hall, A., Scott, J. A., & Shang, H. (2011). Geothermal energy recovery from underground mines. *Renewable and Sustainable Energy Reviews*, *15*(2), 916–924. https://doi.org/https://doi.org/10.1016/j.rser.2010.11.007

Hemami, A., Goulet, S., & Aubertin, M. (1994). Resistance of particulate media to excavation: Application to bucket loading. *International Journal of Surface Mining, Reclamation and Environment*, *8*, 125–129.

Henriksson, E., Söderholm, P., & Wårell, L. (2014). Industrial electricity demand and energy efficiency policy: the case of the Swedish mining industry. *Energy Efficiency*, *7*(3), 477–491. https://doi.org/10.1007/s12053-013-9233-7

Herberich, E., Sikorski, J., & Hothorn, T. (2010). A robust procedure for comparing multiple means under heteroscedasticity in unbalanced designs. *PLoS ONE*, *5*(3). https://doi.org/10.1371/journal.pone.0009788

Hettinger, D., & Lumley, G. (1999). Using data analysis to improve dragline productivity. *Coal Age*, *104*(9), 64–66. https://doi.org/https://doi.org/

Hitachi. (2023). *Mining Equipment (Excavators and Haul Trucks)*. Hitachi Construction Machinery Americas Inc. https://www.hitachicm.us/industry/mining/

Hollander, M., & Wolfe, D. A. (2013). *Nonparametric Statistical Methods*. Wiley. https://books.google.com/books?id=7b05ygEACAAJ

Hu, H., & Kavan, P. (2014). Energy Consumption and Carbon Dioxide Emissions of China’s Non-Metallic Mineral Products Industry: Present State, Prospects and Policy Analysis. *Sustainability*, *6*(11), 8012–8028. https://doi.org/10.3390/su6118012

Iai, M., & Gertsch, L. (2013). Excavation of Lunar Regolith with Large Grains by Rippers for Improved Excavation Efficiency. *Journal of Aerospace Engineering*, *26*(1), 97–104. https://doi.org/10.1061/(ASCE)AS.1943-5525.0000221

Igogo, T., Awuah-Offei, K., Newman, A., Lowder, T., & Engel-Cox, J. (2021). Integrating renewable energy into mining operations: Opportunities, challenges, and enabling approaches. *Applied Energy*, *300*, 117375. https://doi.org/https://doi.org/10.1016/j.apenergy.2021.117375

Jacobs, R., Nel, J., Vawda, Y., Snyman, F., Arndt, R., & Plessis, J.-A. (2018). *Intelligent Mining Delivering Real Value*. Deloitte Touche Tohmatsu Limited. https://www.deloitte.com/global/en/Industries/energy/perspectives/intelligent-mining-deloitte.html

Jammalamadaka, S. R., & Sengupta, A. (2001). *Topics in Circular Statistics*. World Scientific. https://books.google.com/books?id=sKqWMGqQXQkC

Jha, A., & Tukkaraja, P. (2020). Monitoring and assessment of underground climatic conditions using sensors and GIS tools. *International Journal of Mining Science and Technology*, *30*(4), 495–499. https://doi.org/10.1016/j.ijmst.2020.05.010

Jo, B. W., & Khan, R. M. A. (2018). An internet of things system for underground mine air quality pollutant prediction based on azure machine learning. *Sensors (Switzerland)*, *18*(4). https://doi.org/10.3390/s18040930

Kalinowski, P., Długosz, O., & Kamiński, P. (2022). Digital Twin of the Mining Shaft and Hoisting System as an Opportunity to Improve the Management Processes of Shaft Infrastructure Diagnostics and Monitoring. *Data Science, Data Visualization, and Digital Twins*. https://doi.org/10.5772/intechopen.96193

Karpuz, C. (2018). Energy Efficiency of Drilling Operations. In *Green Energy and Technology* (pp. 71–86). https://doi.org/10.1007/978-3-319-54199-0_5

Karpuz, C., Ceylanoğlu, A., & Paşamehmetoğlu, A. G. (1992). An investigation on the influence of depth of cut and blasting on shovel digging performance. *International Journal of Surface Mining, Reclamation and Environment*, *6*(4), 161–167. https://doi.org/10.1080/09208119208944331

Kato, S., Shimizu, K., & Shieh, G. S. (2008). A circular-circular regression model. *Statistica Sinica*, 633–645.

Kecojevic, V., Bogunovic, D., & Kecojevic, V. (2011). Impact of Fill Factor on Dragline Production Rate and Energy Consumption Mınıng engıneerıng Impact of bucket fill factor on dragline production rate and energy consumption. *Mining Engineering*, *63*(8), 48–53. www.miningengineeringmagazine.com

Kecojevic, V., Vukotic, I., & Komljenovic, D. (2014). Production, Consumption and Cost of Energy for Surface Mining of Bituminous Coal. *Mining Engineering*, *66*, 51–57.

Kim, S., & SenGupta, A. (2015). Inverse circular-linear/linear-circular regression. *Communications in Statistics- Theory and Methods*, *44*(22), 4772–4782.

Kim, S., & SenGupta, A. (2018). Regressions involving circular variables: An overview. *Springer Proceedings in Mathematics and Statistics*, *244*, 25–33. https://doi.org/10.1007/978-981-13-1223-6_3

Kirk, T., & Cannon, C. (2020). Sunshine for Mines: A Brighter Vision for Sustainable Resources. *URL Https://Rmi. Org/Sunshine-for-Mines-a-Brighter-Vision-for-Sustainable-Resources/, Accessed August*, *27*, 2020.

Kiziroglou, M. E., Boyle, D. E., Yeatman, E. M., & Cilliers, J. J. (2017). Opportunities for Sensing Systems in Mining. *IEEE Transactions on Industrial Informatics*, *13*(1), 278–286. https://doi.org/10.1109/TII.2016.2636131

Klein, B., Wang, C., & Nadolski, S. (2018). Energy-Efficient Comminution: Best Practices and Future Research Needs. In *Green Energy and Technology* (pp. 197–211). https://doi.org/10.1007/978-3-319-54199-0_11

Komatsu. (2023). *Excavators*. Komatsu Limited. https://www.komatsu.com/products/excavators/

Komljenovic, D., Bogunovic, D., & Kecojevic, V. (2010). Dragline operator performance indicator. *International Journal of Mining, Reclamation and Environment*, *24*(1), 34–43. https://doi.org/10.1080/17480930902778191

Koppelaar, R. H. E. M., & Koppelaar, H. (2016). The Ore Grade and Depth Influence on Copper Energy Inputs. *BioPhysical Economics and Resource Quality*, *1*(2), 11. https://doi.org/10.1007/s41247-016-0012-x

Krishnamoorthy, K., Lu, F., & Mathew, T. (2007). A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models. *Comput. Stat. Data Anal.*, *51*, 5731–5742.

Lanciano, C., & Salvini, R. (2020). Monitoring of strain and temperature in an open pit using brillouin distributed optical fiber sensors. *Sensors (Switzerland)*, *20*(7). https://doi.org/10.3390/s20071924

Lazarević, Ž., Daničić, D., Lazarević, Ž., & Mitrović, S. (2016). *PROACTIVE APPROACH AS CONTRIBUTION TO SYSTEM OF PREVENTING FATIGUE FAILURES OF BUCKET WHEEL EXCAVATOR PROACTIVE APPROACH AS CONTRIBUTION TO SYSTEM OF PREVENTING FATIGUE FAILURES*. https://www.researchgate.net/publication/316748671

Levene, H. (1960). *Robust Test for Equality of Variance*. Stanford University Press.

Levesque, M., Millar, D., & Paraszczak, J. (2014). Energy and mining – the home truths. *Journal of Cleaner Production*, *84*, 233–255. https://doi.org/https://doi.org/10.1016/j.jclepro.2013.12.088

Lezak, S., Cannon, C., & Blank, T. K. (2019). Low-carbon metals for a low-carbon world: a new energy paradigm for mines. *Rocky Mountain Institute. Washington, DC*.

Liu, T., Wei, Y., Song, G., Hu, B., Li, L., Jin, G., Wang, J., Li, Y., Song, C., Shi, Z., Zhao, L., Hu, J., Zhao, W., Hou, M., Li, R., & Wang, J. (2018). Fibre optic sensors for coal mine hazard detection. *Measurement*, *124*, 211–223. https://doi.org/10.1016/J.MEASUREMENT.2018.03.046

Lumley, G., & Lumley, G. (2005). Reducing the Variability in Dragline Operator Performance Reducing the Variability in Dragline Operator Performance Reducing the Variability in Dragline Operator Performance. *Coal Operators' Conference*, 97–106. https://www.researchgate.net/publication/30387716

Lusk, B., & Silva, J. J. (2018). Energy Distribution in the Blast Fragmentation Process. In K. Awuah-Offei (Ed.), *Energy Efficiency in the Minerals Industry: Best Practices and Research Directions* (pp. 11–30). Springer International Publishing. https://doi.org/10.1007/978-3-319-54199-0_2

Maennling, N., & Toledano, P. (2018). The renewable power of the mine. *Available at SSRN 3661616*.

Mansouri, S. S., Kanellakis, C., Kominiak, D., & Nikolakopoulos, G. (2020). Deploying MAVs for autonomous navigation in dark underground mine environments. *Robotics and Autonomous Systems*, *126*, 103472. https://doi.org/10.1016/J.ROBOT.2020.103472

Mardia, K. V. (1976). Linear-Circular Correlation Coefficients and Rhythmometry. *Biometrika*, *63*(2), 403–405. https://doi.org/10.2307/2335637

Mardia, K. V., & Jupp, P. E. (1999). *Directional Statistics*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470316979

McElduff, F., Cortina-Borja, M., Chan, S.-K., & Wade, A. (2010). When t-tests or Wilcoxon-Mann-Whitney tests won't do. *Advances in Physiology Education, 34*(3), 128–133. https://doi.org/10.1152/advan.00017.2010

McKinnon, C. (2022). EMPLOYING CLEANTECH TO INCREASE PRODUCTIVITY AND ENERGY EFFICIENCY AT OPEN-PIT MINES. *SME Annual Conference and Expo 2022*, 304–306.

McKinsey. (2018). Behind the Mining Productivity Upswing: Technology-Enabled Transformation. *McKinsey Metals & Mining*. https://www.mckinsey.com/industries/metals-and-mining/our-insights/behind-the-mining-productivity-upswing-technology-enabled-transformation#/

Minhas, U. I., Naqvi, I. H., Qaisar, S., Ali, K., Shahid, S., & Aslam, M. A. (2018). A WSN for Monitoring and Event Reporting in Underground Mine Environments. *IEEE Systems Journal*, *12*(1), 485–496. https://doi.org/10.1109/JSYST.2016.2644109

Mining Technology. (2022). *Innovations in mining and the future of the industry*. Mining Technology and Ferrit Global Mining Solutions. https://www.mining-technology.com/sponsored/mining-innovations-future-of-mining/

Mishra, P., Singh, U., Pandey, C. M., Mishra, P., & Pandey, G. (2019). Application of student's t-test, analysis of variance, and covariance. *Annals of Cardiac Anaesthesia*, *22*(4), 407–411. https://doi.org/https://doi.org/10.4103/aca.ACA_94_19

Moats, M. S. (2018). Energy Efficiency of Electrowinning. In K. Awuah-Offei (Ed.), *Energy Efficiency in the Minerals Industry: Best Practices and Research Directions* (pp. 213–232). Springer International Publishing. https://doi.org/10.1007/978-3-319-54199-0_12

Mohammad, H. H., Satari, S. Z., & Yusoff, W. N. S. W. (2021). Review on circular-linear regression models. *Journal of Physics: Conference Series, 1988*(1). https://doi.org/10.1088/1742-6596/1988/1/012108

Moridi, M. A., Kawamura, Y., Sharifzadeh, M., Chanda, E. K., Wagner, M., Jang, H., & Okawa, H. (2015). Development of underground mine monitoring and communication system integrated ZigBee and GIS. *International Journal of Mining Science and Technology*, *25*(5), 811–818. https://doi.org/10.1016/J.IJMST.2015.07.017

Motion Metrics. (2023). *ShovelMetrics™ Gen 3 - A complete bucket monitoring solution for all shovels and excavators.* WEIR Motion Metrics. https://www.motionmetrics.com/shovelmetrics-gen3/

Motlogelwa, O. G., & Minnitt, R. C. A. (2013). Optimization of diesel usage at Uitvlugt Mine. *Journal of the Southern African Institute of Mining and Metallurgy, 113,* 345–349.

Muro, T., Tsuchiya, K., & Kohno, K. (2002). Experimental considerations for steady state edge excavation under a constant cutting depth for a mortar specimen using a disk cutter bit. *Journal of Terramechanics*, *39*(3), 143–159. https://doi.org/10.1016/S0022-4898(02)00021-6

Nguyen, H., Choi, Y., Bui, X. N., & Nguyen-Thoi, T. (2020). Predicting blast-induced ground vibration in open-pit mines using vibration sensors and support vector regression-based optimization algorithms. *Sensors (Switzerland)*, *20*(1). https://doi.org/10.3390/s20010132

NIST/SEMATECH. (2013). *Engineering Statistics: e-Handbook of Statistical Methods*. https://doi.org/https://doi.org/10.18434/M32189

Norgate, T., & Haque, N. (2010). Energy and greenhouse gas impacts of mining and mineral processing operations. *Journal of Cleaner Production*, *18*, 266–274.

Odhams, A., Roebuck, R., Lee, Y., Hunt, S., & Cebon, D. (2010). Factors influencing the energy consumption of road freight transport. *Proceedings of The Institution of Mechanical Engineers Part C-Journal of Mechanical Engineering Science - PROC INST MECH ENG C-J MECH E*, *1*, 1–16. https://doi.org/10.1243/09544062JMES2004

Ott, R. L., & Longnecker, M. T. (2015). *An Introduction to Statistical Methods and Data Analysis* (7th ed.). Cengage Learning. https://books.google.com/books?id=VAuyBQAAQBAJ

Pan, W. (2005). Research on condition monitoring and fault diagnosis for hydraulic system of construction machinery. *Doctor Degree Thesis. Xian, Shaanxi, China: Xian Hi-Tech Institute*, 115–144.

Paraszczak, J., & Fytas, K. (2012). Renewable energy sources – A promising opportunity for remote mine sites? *Renewable Energy and Power Quality Journal*, *1*(10), 251–256. https://doi.org/10.24084/repqj10.288

Paredes-Sánchez, J. P., Villicaña-Ortíz, E., & Xiberta-Bernat, J. (2015). Solar water pumping system for water mining environmental control in a slate mine of Spain. *Journal of Cleaner Production*, *87*, 501–504. https://doi.org/https://doi.org/10.1016/j.jclepro.2014.10.047

Park, H. M. (2008). *Univariate Analysis and Normality test Using SAS, Stata, and SPSS*.

Patnayak, S., & Tannant, D. D. (2005). Performance monitoring of electric cable shovels. *International Journal of Surface Mining, Reclamation and Environment*, *19*(4), 276–294. https://doi.org/10.1080/13895260500327912

Patnayak, S., Tannant, D. D., Parsons, I., Del Valle, V., & Wong, J. (2008). Operator and dipper tooth influence on electric shovel performance during oil sands mining. *International Journal of Mining, Reclamation and Environment*, *22*(2), 120–145. https://doi.org/10.1080/17480930701482961

Patnayak, Sibabrata. (2006). *Key performance indicators for electric mining shovels and oil sands diggability.*

Peterson, C. R. (1986). INNOVATION IN MINING TECHNOLOGY. *Mechanical Engineering*, *108*(8), 18–22.

Pewsey, A., Neuhäuser, M., & Ruxton, G. D. (2013). *Circular Statistics in R*. OUP Oxford. https://books.google.com/books?id=qeadAAAAQBAJ

Pishgar, M., Issa, S. F., Sietsema, M., Pratap, P., & Darabi, H. (2021). Redeca: A novel framework to review artificial intelligence and its applications in occupational safety and health. *International Journal of Environmental Research and Public Health*, *18*(13). https://doi.org/10.3390/ijerph18136705

Raj, M. G., Vardhan, H., & Rao, Y. V. (2009). Production optimisation using simulation models in mines: a critical review. *International Journal of Operational Research*, *6*, 330–359.

Ranjan, A., Sahu, H. B., & Misra, P. (2020). Modeling and measurements for wireless communication networks in underground mine environments. *Measurement*, *149*, 106980. https://doi.org/10.1016/J.MEASUREMENT.2019.106980

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen Ways to Look at the Correlation Coefficient. *Source: The American Statistician*, *42*(1), 59–66.

Ronald, C. (2016). *Analysis of variance, design, and regression linear modeling for unbalanced data* (F. Dominici, J. J. Faraway, M. Tanner, & J. Zidek, Eds.; Second Edition). CRC Press Taylor & Francis Group.

Ruff, T. M., & Hession-Kunz, D. (2001). Application of radio-frequency identification systems to collision avoidance in metal/nonmetal mines. *IEEE Transactions on Industry Applications*, *37*(1), 112–116. https://doi.org/10.1109/28.903133

Sadeghi, S., Soltanmohammadlou, N., & Nasirzadeh, F. (2022). Applications of wireless sensor networks to improve occupational safety and health in underground mines. *Journal of Safety Research*, *83*, 8–25. https://doi.org/10.1016/j.jsr.2022.07.016

Sahoo, L., Bandyopadhyay, S., & Banerjee, R. (2018). Benchmarking Energy Consumption of Truck Haulage. In *Green Energy and Technology* (pp. 159–180). https://doi.org/10.1007/978-3-319-54199-0_9

Sánchez, F., & Hartlieb, P. (2020a). Innovation in the Mining Industry: Technological Trends and a Case Study of the Challenges of Disruptive Innovation. *Mining, Metallurgy & Exploration*, *37*(5), 1385–1399. https://doi.org/10.1007/s42461-020-00262-1

Sánchez, F., & Hartlieb, P. (2020b). Innovation in the Mining Industry: Technological Trends and a Case Study of the Challenges of Disruptive Innovation. *Mining, Metallurgy & Exploration*, *37*(5), 1385–1399. https://doi.org/10.1007/s42461-020-00262-1

Sánchez, J. (2018). Solar Energy Applications in Mining: A Case Study. In *Green Energy and Technology* (pp. 273–285). https://doi.org/10.1007/978-3-319-54199-0_15

Sanchidrián, J., Segarra, P., & López, L. (2018). Energy Efficiency in Rock Blasting. In *Green Energy and Technology* (pp. 87–118). https://doi.org/10.1007/978-3-319-54199-0_6

Schlotzhauer, S. D. (2009). *Elementary Statistics Using SAS*. SAS Institute.

Schultz, B. B. (1985). Levene's Test for Relative Variation. *Systematic Zoology*, *34*(4), 449–456. https://doi.org/10.2307/2413207

Scoble, M. J., & Muftuoglu, Y. V. (1984). Derivation of a diggability index for surface mine equipment selection. *Mining Science and Technology*, *1*(4), 305–322. https://doi.org/10.1016/S0167-9031(84)90349-9

Shapiro, S. S., Wilk, M. B., & Chen, H. J. (1968). A Comparative Study of Various Tests for Normality. *Journal of the American Statistical Association*, *63*(324), 1343–1372. https://doi.org/10.1080/01621459.1968.10480932

Sierra Wireless. (2015). *Telemetry Tab*. AirLink Mobility Manager User Guide. https://doc.airvantage.net/amm/reference/maintabs/telemetry_tab/

Singh, S. P., & Narendrula, R. (2006). Factors affecting the productivity of loaders in surface mines. *International Journal of Mining, Reclamation and Environment*, *20*(1), 20–32. https://doi.org/10.1080/13895260500261574

Soofastaei, A. (2016). *Development of an Advanced Data Analytics Model to Improve the Energy Efficiency of Haul Trucks in Surface Mines*. https://doi.org/10.13140/RG.2.2.18587.67365

Soofastaei, A., & Fouladgar, M. (2022a). Energy Efficiency Improvement in Surface Mining. *Latest Research on Energy Recovery*, Ch. 7. https://doi.org/10.5772/intechopen.104262

Soofastaei, A., & Fouladgar, M. (2022b). Improve Energy Efficiency in Surface Mines Using Artificial Intelligence. *Alternative Energies and Efficiency Evaluation*. https://doi.org/10.5772/intechopen.101493

Soofastaei, A., Karimpour, E., Knights, P., & Kizil, M. (2018). Energy-efficient loading and hauling operations. *Green Energy and Technology*, *0*(9783319541983), 121–146. https://doi.org/10.1007/978-3-319-54199-0_7

Soofastaei, A., Knights, P., & Kizil, M. (2017). *Application of Advanced Data Analytics to Improve Haul Trucks Energy Efficiency in Surface Mines*. https://doi.org/10.1201/b22353-12

Srivastava, S. C. (1975). Electronics in Mining Industry. *IETE Journal of Research*, *21*(4), 207–213.

Stephens, M. A. (1969). Tests for the von Mises distribution. *Biometrika*, *56*(1), 149–160. https://doi.org/10.1093/biomet/56.1.149

Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, *69*(347), 730–737. https://doi.org/10.1080/01621459.1974.10480196

Takahashi, H., Hasegawa, M., & Nakano, E. (1998). Analysis on the resistive forces acting on the bucket of a Load-Haul-Dump machine and a wheel loader in the scooping task. *Adv. Robotics*, *13*, 97–114.

Trivedi, V., & Fathi, M. (2021). *Data Analytics Applications in Reducing the Emission Footprint of an Energy System*. https://doi.org/10.1007/978-3-030-72322-4_57-2

Trujillo-Ortiz, A. (2012). *welchanova: Welch's ANOVA test for homogeneity of variances* (1.1.0.0). MATLAB File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/37121-welchanova?s_tid=srchtitle

Trujillo-Ortiz, A., & Hernandez-Walls, R. (2003). *Levenetest: Levene's test for homogeneity of variances* (1.0.0.0). MATLAB File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/3375-levenetest?s_tid=ta_fx_results

U.S. Department of Energy. (2007). *U.S. Mining Industry Energy Bandwidth Study*. https://www.energy.gov/eere/amo/articles/us-mining-industry-energy-bandwidth-study

U.S. Energy Information Administration. (2021). *Annual Energy Outlook*. Retrieved from Https://Www.Eia.Gov/Outlooks/Aeo/.

U.S. Energy Information Administration. (2022). *EIA Annual Energy Outlook 2022*. https://www.eia.gov/energyexplained/use-of-energy/industry.php

U.S. Energy Information Administration. (2023, July 17). *Petroleum & Other Liquids - Gasoline and Diesel Fuel Update*. Independent Statistics and Analysis. https://www.eia.gov/petroleum/gasdiesel/

Van Valen, L. (1978). THE STATISTICS OF VARIATION. *Evolutionary Theory*, 33–43.

Van Valen, L. (2005). CHAPTER 3 - The Statistics of Variation. In B. Hallgrímsson & B. K. Hall (Eds.), *Variation* (pp. 29–47). Academic Press. https://doi.org/https://doi.org/10.1016/B978-012088777-4/50005-3

Verhoeven, R., Willems, E., Harcouët-Menou, V., De Boever, E., Hiddes, L., Veld, P. O., & Demollin, E. (2014). Minewater 2.0 Project in Heerlen the Netherlands: Transformation of a Geothermal Mine Water Pilot Project into a Full Scale Hybrid Sustainable Energy Infrastructure for Heating and Cooling. *Energy Procedia*, *46*, 58–67. https://doi.org/https://doi.org/10.1016/j.egypro.2014.01.158

Vukotic, I. M. (2013). *Evaluation of Rope Shovel Operators in Surface Coal Mining Using Evaluation of Rope Shovel Operators in Surface Coal Mining Using a Multi-Attribute Decision-Making Model a Multi-Attribute Decision-Making Model "Evaluation of Rope Shovel Operators in Surface Coal Mining Using a Multi-Attribute Decision-Making Model."* https://researchrepository.wvu.edu/etd

Watzlaf, G. R., & Ackman, T. E. (2006). Underground Mine Water for Heating and Cooling using Geothermal Heat Pump Systems. *Mine Water and the Environment*, *25*(1), 1–14. https://doi.org/10.1007/s10230-006-0103-9

Wei, B., & Gao, F. (2012). Digging trajectory optimization for a new excavating mechanism of electric mining shovel. *Proceedings of the ASME Design Engineering Technical Conference*, *4*(PARTS A AND B), 1033–1039. https://doi.org/10.1115/DETC2012-70515

Welch, B. L. (1947). The Generalization of Students' Problem When Several Different Population Variances Are Involved. *Biometrika*, *34*(1–2), 28–35. https://doi.org/10.1093/biomet/34.1-2.28

Yaghini, A., Hall, R. A., & Apel, D. (2020). Modeling the influence of electric shovel operator performance on mine productivity. *CIM Journal*, *11*(1), 58–68. https://doi.org/10.1080/19236026.2020.1734405

Yuval, Magen Molho, H., Zivan, O., Broday, D. M., & Raz, R. (2019). Application of a sensor network of low cost optical particle counters for assessing the impact of quarry emissions on its vicinity. *Atmospheric Environment*, *211*, 29–37. https://doi.org/10.1016/J.ATMOSENV.2019.04.054

Zar, J. H. (1941). *Biostatistical Analysis (Pearson Modern Classic)* (D. Lynch, Ed.; Fifth). Pearson Education Inc., 2018. https://doi.org/https://lccn.gov/2017059499

Zhang, L., Zhao, J., Long, P., Wang, L., Qian, L., Lu, F., Song, X., & Manocha, D. (2021). An autonomous excavator system for material loading tasks. *Science Robotics*, *6*(55), eabc3164. https://doi.org/10.1126/scirobotics.abc3164

Zhou, X. H., Gao, S., & Hui, S. L. (1997). Methods for Comparing the Means of Two Independent Log-Normal Samples. *Biometrics*, *53*(3), 1129–1135. https://doi.org/10.2307/2533570

Zhou, X., & Lei, X. (2021). Fault Diagnosis Method of the Construction Machinery Hydraulic System Based on Artificial Intelligence Dynamic Monitoring. *Mobile Information Systems*, *2021*, 1093960. https://doi.org/10.1155/2021/1093960

Zhu, Y., & You, G. (2019). Monitoring System for Coal Mine Safety Based on Wireless Sensor Network; Monitoring System for Coal Mine Safety Based on Wireless Sensor Network. *2019 Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC)*.

VITA

Noah Adekunle Aluko, born in Nigeria, completed his Bachelor of Engineering in Mining Engineering from the Federal University of Technology, Akure, in 2018. Post-graduation, he participated in a year-long National Youth Service until 2020. He was then awarded a full scholarship to pursue his Master's degree and began his studies at the China University of Mining and Technology. However, after a year, due to Covid-19 restrictions, he transitioned to the Missouri University of Science and Technology to continue his studies for a Master's degree in Mining Engineering, with his graduation projected for August 2023. Throughout this period, Noah maintained active memberships with the Society for Mining, Metallurgy, and Exploration (SME), the National Society of Professional Engineers (NSPE), the American Exploration and Mining Association (AEMA), the National Society of Black Engineers (NSBE), and the African Students' Association (ASA). In December 2023, he received his Master of Science degree in Mining Engineering from the Missouri University of Science and Technology.