Repurposing legacy metallurgical data Part I: A move toward dry laboratories and data bank

Yousef Ghorbani a,*, Glen T. Nwaila b, Steven E. Zhang c, Martyn P. Hay d, Lunga C. Bam e, Pratama Istiadi Guntoro a

a Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, SE-97187 Luleå, Sweden
b School of Geosciences, University of the Witwatersrand, Private Bag 3, Wits, 2050, South Africa
c PG Techno Wox, 43 Patrys Avenue, Helikon Park, Randfontein 1759, South Africa
d Eurus Mineral Consultants (EMC), Plettenberg Bay, South Africa
e Department Radiation Science, Necsa, PO Box 582, Pretoria, South Africa

ARTICLE INFO

Keywords:
Data analytics
Dry laboratories
Data bank
Legacy metallurgical data

ABSTRACT

Advancements in modern mineral processing has been driven by technology and fuelled by market economics of supply and demand. Over the last three decades, the demand for various minerals has steadily increased, while the mineral processing industry has seen an unavoidable increase in the treatment of complex ores, continuous decline in plant feed grade and poor plant performance partly due to blending of ores with dissimilar properties. Despite these challenges, production plant data that are routinely generated are usually underutilised. In this contribution and aligned with the direction of the 4th industrial revolution, we highlight the value of legacy metallurgical plant data and the concept of a dry laboratory approach. This study is presented in two parts. In the current paper (Part I), a comprehensive review of the potential for the combination of modern analytical technology with data analytics to generate a new competence for process optimisation are provided. To demonstrate the value of data within the extractive metallurgy discipline, we employ data analytics and simulation to examine gold plant performance and the flotation process in two separate case studies in the second paper (Part II). This was done with the aim of showcasing relevant plant data insights, and extract parameters that should be targeted for plant design and performance optimisation. We identify several promising technologies that integrate well with existing mineral processing plants and testing laboratories to exploit the concept of a dry laboratory, in order to enhance pre-existing mineral processing chains. It also sets the passage in terms of the value of innovative analysis of existing and simulation data as part of the new world of data analytics. Using data- and technology-driven initiatives, we propose the establishment of dry laboratories and data banks to ultimately leverage integrated data, analytics and process simulation for effective plant design and improved performance.

1. Introduction

The mineral processing value chain is complex and consists of several different sequential processes and operations. Traditionally, these processes and operations were conducted in gated stages, reflecting the compartmentalised nature of the mining industry. This can potentially complicate the plant optimisation process, as local optimisation in one operation does not guarantee a universal solution to other sections of the mineral processing value chain (Ding et al., 2017). This entails that if each operation works independently (silo-approach), the optimisation of the whole value chain cannot be performed; there is a need for an integrated approach to achieve mineral processing value chain-wide optimisation.

The emergence of integrated approaches in mineral processing industries such as geometallurgy (Aasly and Ellefmo, 2014; Dominy et al., 2018; Lishchuk et al., 2020; Lund et al., 2015) and process mineralogy (Becker et al., 2016; Lotter, 2011) is largely based on breaking the silo-approach to optimise the whole production process and discover new synergies. Using an integrated approach, the optimisation of ore beneficiation process can already be achieved in upstream processes through,
for example in the approach known as “mine-to-mill optimization” blasting is controlled in mining to increase comminution efficiency (Chi et al., 1996; Yu et al., 2011; Chai et al., 2014) or the ore blending selectively is performed to produce favourable feed for the downstream processes (Aasly and Ellefmo, 2014; Lund and Lamberg, 2014; Philander and Rozendaal, 2014; Nwaila et al., 2019).

These approaches require good communication as well as access, where historical and operational data can be easily utilised in each optimisation stage of the production process. The availability of operational data across the whole plant is already a common practice in chemical process industries (Marchetti et al., 2014). By continuously updating the process data to reflect the current condition in the plant, the process can be optimised in real-time, commonly referred to as Real-Time Optimisation (RTO) (Taha and Rashid Khan, 2011). In such a system, process parameters are continuously measured until a steady state is detected. When a steady state is achieved, the online process model is updated to simulate the current plant operation. Process optimisation is then conducted using the updated model. New optimised operating parameters are then obtained and implemented in the plant through distributed control systems (DCS). This cycle can be repeated again after the optimised operating parameters have been implemented and the steady-state condition has been achieved. This process implements a feedback control system at an operating level.

The challenge in implementing such optimization procedures in mineral processing plants is associated with the difficulty of establishing quantitative process models that describe the operations (King, 2001). The majority of models established in the mineral processing industry are limited in their application to their respective case studies, as each plant has its own unique features and raw materials. Moreover, Ding et al. (2017) argued that in a single plant, process optimisation is often only approached for a single unit operation; there are comparatively few cases where global optimisation for the entire plant is achieved.

Furthermore, the development of digitalized, online plant-monitoring systems for the mineral processing industry is generally limited to that of online chemical analysis using X-ray Fluorescence (XRF) (Nakhaei et al., 2012a). While such chemical analysis is useful, it is often inadequate in explaining the complex mineralogy of the streams and their interconnection with process behaviour, which drove further development to mineralogical-based concept of geometallurgy and process mineralogy in recent years. Development of geometallurgical tools such as element to mineral conversion (Lund et al., 2013) are mainly driven by the need of rapid acquisition of mineralogical information in mineral processing plant operations, allowing access to new types of data and additional insights to the process behaviour.

With respects to raw material and ore characterization, mineralogy and rock textures are still largely qualitatively analysed by a trained human operator using a petrographic microscopy and/or hand lens. This poses challenges as it means that historical ore data is difficult to utilize especially in modern data analytics and process simulation, much of which needs quantitative data. With the rapid development of ore characterization and tools such as X-ray Diffraction (XRD), Automated Mineralogy, as well as 3D X-ray Computed Tomography (XCT) for quantitative mineralogical and textural analysis of the ores, it is paramount to establish data banks to integrate highly quantitative data and thereby maximising its value for simulation and optimisation of plant performance.

This study is structured in two parts. In the first paper (Part 1), a review of recent technological advances in the metallurgical discipline covering the full-spectrum of the mineral processing value chain is provided. Part 1 also examines the potential for the integration of modern analytical technology with data analytics to create new process optimisation opportunities. We assess the value of operational and legacy data in metallurgical plant for global production optimization using an integrated and holistic approach through evaluation of plant operational data, ore characteristics data, as well as legacy / historical data from other sections involved in the production value chain, such as mining, geology, and metallurgy. The availability of such data would enable the shift to dry laboratory* environments as a future-oriented approach, where the optimisation is performed with a model-based simulation of the real-life operation, similar to that of RTO. The second paper (Part 2) delves into two case studies, setting the scene in terms of the value of novel just-in-time (low-latency and near real-time) data analytics and simulation as part of the new world of technology-driven mineral processing.

2. The value of legacy metallurgical data in the industry

Metallurgical operations globally generate large datasets (i.e. big data) on a daily basis (McCoy and Auret, 2019). Technology firms such as Amazon and Google leverages legacy and real-time data, including big data by judicious applications of data-mining through artificial intelligence, machine-learning and deep-learning algorithms to gather, process and transform data, and to automatically discover patterns and extract insights. Such insights are used to drive business growth, market-share and mindshare. In metallurgical plants, the situation pertaining to the storage and use of data is decidedly less modern. A bulk of historical and current plant operational data (“operational data”) are stored in paper reports, digital spreadsheets and/or software systems, which are then filed off for compliance purposes. Over the last few years, there has been a proliferation of initiatives that employ various combinations of equipment, data acquisition, and processing techniques, to improve plant performance and metal recovery (Makokha and Moys, 2012; Jahedsaravani et al., 2016, 2014; Massinaei and Doostmohammadi, 2010; Nakhaei et al., 2012b). The most successful solutions to date have come from the deployment of online sensors and central monitoring system (CMS), which collects real-time data that enables just-in-time process visualisation and rapid management response (Jámsí-Jounela, 2019). Another area of analytics in mineral processing plants has focused on predictive analytics for planned plant maintenance, which is a technique that uses data to understand equipment degradation and predict failures (Botha et al., 2018). However, numerous difficulties have been encountered, even with these innovative data analytics approaches. Metallurgical data acquired in the actual processing plant suffers from quality, quantity and consistency issues. Furthermore, in some metallurgical operations, metallurgical accounting is combined with metal accounting and data is not collected at regular intervals, thus creating additional challenges for analytics (Gaylard et al., 2009; Ghorbani et al., 2020).

In a fast-evolving technological era, legacy metallurgical data is often regarded by plant engineers as inferior compared with the newly acquired, since technology itself drives rapid progress in in-situ monitoring instrumentation. In this context, the term “legacy data” refers to metallurgical and metal accounting data stored in an old or obsolete format, or in a manner that requires substantial data pre-processing and

---

1 Steady state here means that when the variable has approached the set point value.

2 A dry lab is a laboratory where computational or applied mathematical analyses are done on a computer-generated model to simulate a phenomenon in the physical realm.

3 Metallurgical accounting (primary accounting) refers to a system whereby selected process data are collected from various sources across an entire process and transformed into a coherent metallurgical report to meet specified operational and reporting requirements. Metallurgical accounting (secondary accounting) refers to a system whereby selected process data (pertaining to metals of economic interest) is collected from various sources including mass measurement and analysis and transformed into a coherent report format that is delivered in a timely fashion in order to meet specified reporting requirements (Gaylard et al., 2009; Ghorbani et al., 2020).
cleansing in preparation for data analytics. We use the terms “operational / current production data” interchangeably to refer to up-to-date actual data from the production process. Routine operational data contains valuable information regarding relationships between common feed characteristics and performance indicators. Furthermore, these relationships provide indirect knowledge of the structure and/or composition of the feed sources, such that the resulting knowledge is already appropriate to guide the optimisation of the operation. The existing operational data that is generated on a daily basis could be combined with automatable analytics to produce data-driven insights and actionable knowledge to drive continuous improvement in plant performance (Jahedsaravani et al., 2017). Therefore, operational data is an essential mining asset that will only prove to be more valuable as the industry transitions towards a data-rich and increasingly data-driven type of operational feedback.

We propose the establishment of dry laboratories, a focus of which would be on the analysis of legacy (and modern) data to extract actionable knowledge that span several timescales. The short-term knowledge can be leveraged to provide just-in-time feedback to existing processing plants and the long-term knowledge can be used to guide the design and refinement of new plants. The new knowledge obtained from legacy data may benefit future metalurgical operations by providing better metrics to monitor and assess complex ore treatments and enable cost reductions in the operation and design of metallurgical plants. Successful applications of the above approach could enable mineral processing engineers to rapidly design and refine processes utilised within new and existing metallurgical plants, while providing a foundation to enable processing plants to leverage a range of data analytics algorithms that are on-par with those utilised by Amazon and Google.

3. Driving factors moving towards dry laboratories and data banks

3.1. Availability of the state of art characterisation tools

3.1.1. Automated mineralogy

It has been long realised that there exists a connection between the mineralogy of a feed and its subsequent metallurgical performance (Gaudin, 1939; Petruk, 1976). This is further advanced with the development of sampling theory (Gy, 1979) as well as the concept of statistical uncertainties entailed with the sampling procedures for particulate materials. Subsequently, automated quantitative mineralogical analyses together with the Scanning Electron Microscopy and Energy Dispersive X-ray Spectroscopy (SEM-EDS) instruments were developed. Such instruments enable rapid and automated measurement of mineralogy, particle morphology, as well as texture of ore samples (Fandrich et al., 2007; Gottlieb et al., 2000; Gu et al., 2014; Sutherland, 2007; Sutherland and Gottlieb, 1991). All of these developments ultimately form an integrated approach of process mineralogy (Becker et al., 2016; Lotter, 2011).

Many case studies demonstrated the use of integrated approaches for optimisation of metallurgical and beneficiation processes. The use of mineralogical information in mineral processing operations has been studied by Lotter (2011) and Whitman et al. (2016), with both studies demonstrating the increase of metal recovery after incorporating mineralogical and textural information in the flow sheeting and optimisation process. Other studies include, for example, utilisation of mineralogical and textural information of iron ores to optimize the subsequent downstream operations such as sintering and hydrocyclone processes (Domskoi et al., 2019). Utilization of mineralogical information can be extended as well to mine planning and ore body development, in which the mineralogical information of the ore blocks can be used to predict the processing behaviour of the block. The ore blocks can then be classified into both geological and metallurgical types, thereby adding information in the mine planning for maximizing the whole mine-to-mill value chain (Gottlieb, 2008).

It is, however, worth mentioning that automated mineralogy measurements have some issues such as related to sampling and statistical errors (Lätti and Adair, 2001; Ueda et al., 2016). Additionally, while automated mineralogy provides many valuable information about the sample, other methods such as quantitative XRd can provide greater details about the gangue minerals (Parian et al., 2015) or SATMAGAN analysis that can provide more details on the presence of magnetite minerals (Lund et al., 2015). The vast amount of data generated by automated mineralogical instruments necessitates the establishment of a data repository, to facilitate easy accessibility for various divisions in a mining operation. Additionally, the establishment of a highly-structured and centralized data repository, such as a data warehouse, would facilitate integrated data and information intelligibility, increase confidence in analysis and forecasting, improve data quality and deliver high-quality business intelligence. The establishment of a structured data repository with rigorous data governance would enable the cross-correlation and validation of mineralogical data generated from various different analytical tools, thereby potentially decreasing the statistical uncertainty in the data and thus in the derived information. For example, there are a couple of techniques available to minimize the statistical uncertainty when using mineralogical information for mass balancing and modelling purposes, typically by cross-correlating the mineralogical data to XRF chemical analysis of the sample (Alves dos Santos and Galery, 2018; Lamberg and Vianna, 2007).

The potential of the data repository could also be expanded to include experimental databases. Database systems such as LIMS (Laboratory Information Management System) which is traditionally used as a means of managing assays analysis could potentially be extended to include mineralogical analysis as well as metallurgical testing data. Metallurgical laboratory testing is often necessary to correlate mineralogy-to-process performance in a process-mineralogical approach (Lotter, 2011). The reproducibility of laboratory tests is critical and the ease of replication is desirable, such that the testing methodology is maximally generalisable, as well as to validate the original experimental results. This also allows for continuous improvements and updates of the simulation to reflect ongoing changes in the plant. Combined with the mineralogical data repository, process simulation could also be linked to the input data, avoiding inconsistencies in the input parameters to ensure reproducibility of the simulation. Similarly, mineralogy and mineral liberation data obtained from automated mineralogy can be used to model beneficiation processes such as flotation and magnetic separation (Alves dos Santos, 2018; Lotter et al., 2003; Parian et al., 2018, 2016) for process optimisation. The idea of the creation of an experimental database has been coined by Perianayagam et al., (2019). Such an idea was originally intended for software experiments, as reconstructing such experiments are often problematic due to various reasons, such as missing input libraries, compatibility issue with target platform, and inconsistent command-line arguments. Nevertheless, such a data repository is also relevant and applicable to data generated by automated and manual instrumentation, and by extension, process modelling and simulation.

3.1.2. X-ray tomography

Recent decades have shown promising progress for the use of 3D X-ray Computed Tomography (XCT) for mineralogical characterisation. XCT is a non-destructive technique that is capable of producing rapid 3D mineralogical and spatial information of ore samples (Barn et al., 2016; Ghorbani et al., 2011; Coshell et al., 1991). This is possible because the technique registers the linear attenuation coefficient information of each mineral present within the sample. The linear attenuation coefficient is the sum of the scattering and absorption coefficients of a material, which depend on the internal structure and density of the material and is often unique for each mineral at a specific X-ray energy. In some cases where minerals share similar linear attenuation coefficients at a specific X-ray
energy, discrimination between the minerals can be affected. However, for mineral processing, many common minerals are identifiable through XCT, such as gold or platinum and sulphide mineral groups. Linear attenuation coefficients for some common minerals such as native gold, platinum, sulphide mineral groups (e.g., pyrite, chalcopyrite, and galena) and silicate minerals (e.g., quartz) are shown in Fig. 1.

The differences in attenuation of various minerals are exploited in XCT analysis for grade-data acquisition in order to preselect samples with desirable amounts of a target mineral. The high linear attenuation coefficient of a target mineral (e.g. gold) provide guidance in the selection of optimal sample size and scanning parameters in order to maximise scanning resolution and minimise the impact of beam hardening. The combination of optimal sample size and scanning parameters facilitates rapid data acquisition (Bam et al., 2016; 2019) and real-time feedback.

It is worth mentioning that real-time feedback to the level employed in the chemical industry requires high performance computers to enable fast and automatic reconstruction of 3D images during scanning. The reconstructed data with clear discrimination of target mineral grains can be easily quantified and used to determine grain and particle size distributions (Evans et al., 2015; Le Roux et al., 2015; Rozendaal et al., 2018), which are important for optimizing comminution and the optimal feed grade. The 3D data can also be used for quantitative textural analysis, which is especially relevant for automated drill core recognition and classification (Guntoro et al., 2020a,b; Jardine et al., 2018; Voigt et al., 2019).

XCT analysis could also be coupled with in-situ experiments such as leaching (Dobson et al., 2017; Ghorbani et al., 2013a) and breakage tests (Alikarami et al., 2015). The reconstructed data from in-situ leaching tests can be further utilised to track mineral leaching progress (Fagen-Endres et al., 2017; Lin et al., 2016; Ghorbani et al., 2013a) of various mineral particles at different time intervals. A similar analysis could also be adopted for in-situ breakage tests, which could provide crucial information on crack propagation in ore particles for breakage and liberation modelling (Garcia et al., 2009; Xu et al., 2013; Ghorbani et al., 2013b). These coupled in-situ XCT analyses could provide valuable time-series data of metallurgical tests, which could help in building models for process optimisation.

Furthermore, XCT can also be utilised as a complimentary tool in ore characterisation, especially with synchrotron sources. XCT-based complimentary tools such as X-ray Diffraction Tomography (XRD-CT) and X-ray Fluorescence Tomography (XRF-CT) have found applications for analysis of geological materials (Artioli et al., 2010; Laforce et al., 2017; Suuronen and Sayab, 2018; Takahashi and Sugiyama, 2019). Otherwise, recent developments have also made possible the use of common laboratory polychromatic-source XCT in different modes, such as diffraction-contrast and phase-contrast modes, which could be useful for mineral classification (King et al., 2014; Olivo and Castelli, 2014; Viermetz et al., 2018).

XCT is an important tool and can add value to mineral processing for its generation of large volumes of useful data that can be used in a dry laboratory approach for process design. The graphic nature of raw XCT data is also highly conducive to machine vision-based approaches to classification and predictive analytics, such as improvements in particle characterisation, phase discrimination and process modelling. Additionally, the developments of coupled analysis of various analytical tools as well as in-situ experiments with XCT systems further enhance the value of data generated from XCT systems. However, for application in mining industries, the current laboratory XCT systems need to be adapted in terms of design similarly to the ones used in the forest industry (Wei et al., 2011). This will allow a large number of samples to be scanned simultaneously and large amounts of data to be generated rapidly for continuous production optimisation.

3.1.3. Positron emission particle tracking

Positron emission particle tracking (PEPT) was developed at the University of Birmingham (UK) to track the particle flow in dense and opaque engineering systems (Hawkesworth et al., 1986; Parker et al., 1993). A full description of PEPT is presented in Leadbeater (2009) and Leadbeater et al. (2012). PEPT is based on tracking an individual particle tracer labelled with a positron-emitting radionuclide that decays by positron emission (Parker et al., 1997). Each radioactive decay leads to

---

**Fig. 1.** Attenuation. Linear attenuation coefficients of various minerals. Clear differences are observed between high density minerals (e.g. gold, platinum), sulphide minerals (e.g. pyrite, chalcopyrite), and silicate minerals (e.g. quartz, biotite).
the emission of a positron, which quickly annihilates with an electron, and nearly always creates a pair of gamma rays. They are emitted nearly exactly back-to-back due to conservation of linear momentum, so that simultaneous detection of both gamma rays provides a geometrical constraint on the tracer’s position in the form of a line through a point in close proximity to the tracer’s position at the time of positron emission. A precise location can then be gained by triangulation of a small number of such events. There are different methods to label particles’ direct activation, ion exchange (Fan et al., 2006a) and surface modification (Fan et al., 2006b). PEPT is the only existing technique capable of imaging complex flow, mixing and comminution (grinding) processes within an opaque industrial equipment. Prior to PEPT, much of the understanding and optimisation of industrial flow and mixing processes were obtained through mathematical and computational models that were impossible to directly validate. PEPT allows a paradigm shift in process design and multi-scale modelling. Although PEPT might be a useable technique for generating plant-related data in the near future, it shows significant promise as a research tool in the mineral processing industry (Table 1).

3.2. Digitalisation and Online-Monitoring systems

Data has always been the starting point of any mining operation and mineral processing plant. Many mining operations and mineral-processing plants are currently using online sensors in their equipment for process visualisation, quality management, process events and alerting management (Kadlec et al., 2011; 2009; Lin et al., 2007; Janusz et al., 2017; 2016; Souza et al., 2016). Acquisition of data is commonly through a supervisory control and data acquisition (SCADA) system and monitored in a central monitoring system (CMS) or control room. Online process analytics, supported by plant specific mass flow models, enable accurate and timely reporting of metal balances. In addition, online sensors can lead to: (a) optimised stockpile management and ore blending; (b) improved milling performance and reagent use; and (c) improved plant recovery (Mazzour et al., 2003; Wang et al., 2016; Dindarloo and Siami-Irdemoosa, 2017). More recently, online reagent analysers have been proven to produce measurements with accuracy levels that are comparable to those obtained in a laboratory. Application of online analysers include on-site applications near the point of sampling, which provides results within a few minutes. This prevents long holding times that make it difficult to accurately analyse substances for their concentrations of reagents and metals in the laboratory. Moreover, the near-instantaneous availability of the results means that process plant operators are equipped to make swift decisions rather than being forced to delay their operations for hours or days. The ability to carry out continuous monitoring also means that the process allows the state of the environment, as well as the mining operations, to be visualised on a dynamic rather than a static, snapshot-like basis, as is the case when grab sampling is used. This improves the reliability of metal accounting, enables further process optimisation and early detection of anomalous events in the process.

3.3. Smart Multi-filtering system to identify specific reagents

The rise of modern artificial-intelligence methods has the latent potential to significantly change and boost the role of computers and data in science and engineering (Yang et al., 2010). The combination of big data and artificial intelligence (AI) has been referred to as the “fourth paradigm of science” and the “fourth industrial revolution” (Schwab, 2013). Computational chemistry has become progressively predictive, with a bustle in applications as wide-ranging as catalyst development, materials discovery, and computer-assisted medicine design. High-throughput computational screening has become ubiquitous, providing scientists with the capability to compute the properties of thousands of compounds as part of a single study. Chemicals are used in nearly all the steps in mineral processing and extractive metallurgy, such as grinding (wet and dry grinding aids), flotation, hydrometallurgy, solid–liquid separation, tailings treatment, and materials handling (Prasad, 1992; Lewis, 1992; Hutton-Ashkenney et al., 2015). The use of chemicals in the mining industry continues to grow in volume and complexity as operators look to improve recovery and extraction techniques and maximize yields from complex ores. As ores become more challenging and process water becomes more impure, requirements grow with respect to environmental and health protection, meaning that more tailored and inventive chemical solutions will be needed.

Applying a smart tool with a multi-filtering system could help to formalize a decision-making process framework dedicated to the sustainable selection of reagents in an integrated scientific and industrial context. It would help to identify mineral structure- and process-specific reagents. Such a smart tool could be supported by data mining, which discovers patterns in large datasets using methods such as artificial intelligence, machine learning, statistics, and database systems. This leads to the integration of computational molecular structure and reagent chemical interaction analysis with mineralogy, crystallography and/or the physico-chemical condition of the relevant reactions in order to identify the most effective reagents. This approach will certainly help the mining industry to maximise the performance of their operations through optimised reagent selection, superior application expertise, and a constant stream of innovative products and technologies.

3.4. Data bank laboratory establishment

The entire history of natural resource exploitation is rich with data. The operational and financial aspects of the mining industry alone have necessitated the collection, analysis and storage of data. Data resulting from the operational side of the business documents details of the operations, such as characterisation of the ores at the bulk- and micro-scales, tonnages processed, dates and worker shift productivity. Such data is very useful for the evaluation of operational efficiency and to identify any potential issues that lowers productivity and/or efficiency. This type of information can be immediately turned into business intelligence and allows for process refinement and strategic business planning.

However, there are several challenges in the mining industry regarding data in context with modern data standards. Since competition forces private mining businesses to shelter their data, there is little in terms of data standardisation across businesses. Therefore, data quality dimensions, such as accuracy and precision, consistency and completeness can be wholly missing outside of process-related quality control and assurance. This is a major issue that requires an integrated approach or standard to be taken across the mining industry, to at least develop a rigorous standard that is usable across the industry. Such a standard is unlikely to originate from a single business or company.

Another major issue is non-centralised storage of a variety of data. Within only the operational-side of the mining industry, data is generated from analytical instruments, experiments, qualitative and quantitative geology, geostatistics and plant operations. While there are clear links between these areas, there is typically no centralised data repository. Some advanced mining operations are moving towards data warehousing or data lakes, but the bulk of operations in the industry creates a large variety of data that are essentially siloed. This issue complicates data analysis and makes tracking of business processes and performance of each segment of the pipeline difficult and cumbersome. Without a centralised store of at least partially structured data, an integrated, easy-to-use business intelligence tool would be impossible.

This paper does not propose a solution to an industry-wide problem as this issue is monumental and requires the collective participation of many private industries to solve. Instead, we observe that within the laboratory setting, we can replicate many standard industrial processes and source common materials for small-scale and highly reproducible investigations that simulate various processes that exist within the mining industry, such as gold extraction. This is a highly controlled
environment where the data collected can be ensured in their quality and management. Data collected within this environment can be used to gain insights into the processes employed by the businesses and offer low-cost, high-speed feedback to the businesses that is typical of laboratory services, and allow evidence-based business intelligence to be utilised at the business scale.

Broad and deep knowledge can be extracted from an abundance of aggregated data and its implications on extraction and processing pipeline design are profound. Such knowledge is generally impossible to produce through siloed data from a single plant or simulation. Indeed, we showcase knowledge generation from aggregated simulation data on the second part (Part II) of this paper. Thus, it is important to aggregate data from multiple sources in order to produce generalisable information that transcends the boundaries of a particular ore, plant or business. Aggregating data in the form of a data bank is an ideal approach.

The amount of non-public domain data that is essentially inaccess-ible outside of the industry is vast. This issue reflects the nature of siloed operations typically conducted within the private industry sector. With changes in the industry, including closures and personnel replacements, data is gradually lost or inadvertently destroyed, especially on exhausted ores. This reinforces the notion of sound data governance and the need for independent laboratories with a significant data generation and management component, such as dry laboratories, to ensure that within the competitive private industry setting, important knowledge that can provide a universal benefit is not lost due to the tragedy of these above effects.

The dry lab approach does not aim to completely replace wet lab and analytical laboratory test work. There is still a need for rigorous laboratory work with a clear focus and goal to mend knowledge gaps that would be otherwise unfeasible or undesirable to address at the plant level. Laboratories⁴ can rapidly replicate plant processes to augment existing data, explore nuances such as ore reaction kinetics that are sometimes impossible to study at the plant level, and provide a neutral, well-managed and data generation source of high quality, free of silos and intended for knowledge production through data analytics. The effective use of concepts driven by big data starts with recording the right data. The Pareto principle states that a minimum number of key variables can adequately describe any system and these are all that are necessary in order to make a business decision such as the design and configuration of a flotation plant and its revenue-generating capability. In flotation, there are > 40 variables that affect the process in varying degrees. Including all variables into a model unnecessarily complicates analysis such that it becomes unusable to its users, as it becomes difficult to establish direct and simple correlations between measured variables and plant performance. These variables can be dimensionally reduced to the extent that over 90% of flotation performance variability can be accounted for by 8 kinetic parameters obtained from a laboratory rate test (Fig. 2). Each kinetic parameter has a physical meaning in a production plant and when incorporated into a model demonstrates the same patterns of behaviour as the real system. The key to using a database effectively is the construction of a model that is simple and effective and is only complex enough to meet these requirements.

4. Process simulation data and its application in dry labs

The potential application of dry laboratories in mineral processing simulation is extensive, as the simulation could include not only operational and technical data, but also environmental data (Pell et al., 2019; Segura-Salazar et al., 2019) and techno-economic data (Khalesi et al., 2015). With the wide availability of supporting data for such simulation, a holistic and integrated approach in optimizing mineral and metallurgical process simulation could be realized.

Process simulation data can be beneficial in the early stages of the project, especially during commissioning and upscaling. Seppälä et al. (2014) demonstrated a combination of laboratory experimental works as well as steady-state process-simulation data in commissioning a mini-pilot scale flotation line. The pilot-scale flotation and its process simulation were then used to study a full-scale flotation process of Pyhäsalmi mine (Veijola, 2014), which allows a detailed evaluation of various operating parameters and their effect to the flotation process. Other examples of upscaling projects with the aid of process simulation data and laboratory works include the scaling up of vertical mill (Mazzinghy et al., 2014). The vertical mill has been known to be more energy effective than standard tube-ball mill, and the process simulation helped to evaluate how the energy-efficiency of vertical mills scales with size.

The design practice of the mining and metallurgical industry is typically separated between the technical and economic analysis, i.e. the technical flowsheet development and feasibility study are done separately (Khalesi et al., 2015). The ideal practice would be an integrated process simulation and cost estimation approach using both technical and economic constraints to achieve the best compromise between performance and cost. Such practice is widespread in chemical engineering industries, but it is comparatively rare in the mineral processing industry. Decision-making in the early stages of the project can be assisted using integrated process simulation that uses technical data as well as cost variables, for example, the decision about the optimum number of leaching tanks required has been found to be affected by the interest rate (Khalesi et al., 2015), while the use of operational cost variables can help in designing the capacity of flotation cells (Arfania et al., 2017).

Furthermore, as the mining industry is constantly pushed towards a more sustainable approach, the inclusion of environmental and Life Cycle Analysis (LCA) in the process design stage are also inevitable. A recent review (Segura-Salazar et al., 2019) has demonstrated that LCA studies in mining operations is often separated from operational process simulation. Such integration could be of great value especially in the early stage of mining projects, where decision-making can have a big impact on the mining operations and subsequently their environmental impacts. While environmental improvements can be made during the lifetime of a mining operation as more data becomes available (Pell et al., 2019), the most potential improvement is in the early stages of the project when different scenarios can be evaluated. Therefore, it is critical that process simulation must be integrated with LCA, such that environmental impacts can be minimised with minimum interference to the process performance. Examples of such approaches include choosing a suitable energy source with respects to its reliability and global

### Table 1

<table>
<thead>
<tr>
<th>Application</th>
<th>Particle behaviour in mills</th>
<th>Flow pattern of slurry in a spiral concentrator</th>
<th>Laboratory batch jig</th>
<th>Hydrocyclones</th>
<th>Flotation cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>References</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ibbosa, et al., 2011; Kallon et al., 2011; Jayasundara et al., 2011; Govender et al., 2013; Morrison et al., 2016; Parker, 2017; de Klerk et al., 2019</td>
<td>Boucher et al., 2014; Boucher et al., 2016; Boucher, 2017</td>
<td>Roux and Naudé, 2014.</td>
<td>Chan et al., 2009; Fonnes, 2011; Chang et al., 2011a,b; Chang et al., 2012; Chang and Hoffmann, 2015; Hoffmann et al., 2019</td>
<td>Waters et al., 2008; 2009</td>
<td></td>
</tr>
</tbody>
</table>

⁴ The laboratories here are mainly mean data exploration and modelling. It is not included LIMS in its traditional meaning, which are essential operational labs for analysing assays to establish product quality along with the mineral processing plants. However, as described above the extended version of LIMS could also significantly benefit from such an approach.
warming impact in rare earth elements (REE) processing plant (Pell et al., 2019), evaluation of different processing routes for spodumene beneficiation for lithium products applicable to electric vehicle batteries (Oliazadeh et al., 2018), as well as optimizing iron ore blending strategies to increase the energy efficiency of the comminution circuits while meeting product quality requirements (Mkurazhizha, 2018). In addition, sustainable planning for mining projects should re-examine the traditional maximisation of supply-chain efficiency to account for a desirability for increased supply-chain robustness to survive global disruptions in hindsight of the global COVID-19 pandemic.

5. Synchronised metallurgical plant data acquisition

The vast amount of data acquired routinely in a metallurgical plant can contribute differently in optimising the performance of the whole plant. Synchronised acquisition of these data in conjunction with a databank or perhaps in some cases, a data warehouse establishment is required to achieve global optimisation of the metallurgical plant. A cloud-based platform for monitoring operational data of various unit operations in a metallurgical plant has been proposed (Xu et al., 2018). This platform allows real-time data collection for continuous optimization and troubleshooting of the plant. Such a platform could also be integrated with other business information systems, such as enterprise resource planning (logistics, warehousing, and inventory management) as well as planning and scheduling (Spackova et al., 2019).

Synchronised acquisition of plant data can also be useful for predictive maintenance in metallurgical plants, which seeks to predict points and modes of failure and promotes proactive maintenance (Fernandes et al., 2018). Predictive maintenance has become more popular compared to traditional corrective maintenance, which failures are detected and diagnosed, and thereafter reactively corrected (Mobley, 2002). Combined with a big-data approach, predictive maintenance could be used not only for failure diagnosis but also to suggest a possible course of action that could further optimise the plant’s existing process in terms of performance, reliability and energy consumption.

The use of metallurgical plant data for process simulation as well as metal accounting often requires data reconciliation to improve the data accuracy and reduce error proliferation (e.g. through dependence and derivation) in the acquired data. Data reconciliation in some metallurgical plants is difficult when metal content in the ore is very low and heterogeneously distributed in the plants, as cases with gold processing plants (de Andrade Lima, 2006). In such cases, the use of thermodynamic as well as mass and energy balance constrains in the data reconciliation process is required, such that the acquired plant data is suitable for process simulation and optimization as illustrated in Fig. 3.

Furthermore, the understanding of what constitutes as operational data has also become widened over the years, as data acquisition techniques becomes more modernized and varied. Image data acquisition using machine vision technology has been growing over the last decades and has found application in mineral processing and metallurgical plants. Applications of machine vision in ore sorting and grade estimation (Perez et al., 2011; Tessier et al., 2007), froth texture analysis (Kistner et al., 2013), online particle-size analysis (Ko and Shang, 2011), as well as quality control of steel surfaces (Bharati et al., 2004) are known. Most of these researchers devoted their work in the feature extraction of image data for a wide range of applications in mineral processing and metallurgical plants. However, latest deep-learning techniques that makes autonomous driving possible have yet to be incorporated into mineral processing, although the underlying techniques should adapt well to online monitoring systems with highly visual data (e.g. XCT and PEPT). A detailed discussion have been provided at Section 3.1 (Data-based modelling in minerals processing) in the latest review paper on Machine learning applications in minerals processing by McCoy and Auret (2019).

All of the approaches described herein detailed how data acquisition, processing, and reconciliation is the first key step in achieving optimised plant performance. The value of the data acquired from different sections of the plant can be maximised with robust process simulation in a dry lab environment, thereby creating an integrated approach in global optimisation of plant performance leveraging known and emerging technological developments.

6. Summary and outlook

Technological evolutions in the mineral processing industry including the use of automated mineralogy techniques, best-reagent mixtures and data analytics have the potential to provide key process insights, thus enabling effective plant design and performance optimisation. Common to all these exciting developments is the reliance and/or generation of data. An adequately integrated approach consisting of
existing and new technologies that adopts a solid data pipeline from generation to analytics can enable timelier (e.g. just-in-time), more accurate and diverse feedback relative to the current state. Therefore, the data pipeline and the adequacy of its integration will likely become crucial factors controlling the outcome of the adoption of the 4th industrial revolution concepts in the mineral processing industry. The mineral processing industry is universally limited by modern economic margins and impacted by increasingly mineralogically complex ores, which necessitate increasingly optimised and adaptable plant processes. Optimisation of plant processes should be built on a baseline foundation of detailed historical findings that are generalisable and dependable, and further customized and adapted dynamically at the operation level by insights derived through analytics from operational and laboratory data.

Together, operational and historical data covers a range from detailed to large-scaled plant data. These types of data can be considered to be the end-members of a spectrum of data, in the sense that operational data captures the short-term reality and can provide timely feedback to the operation, and historical data is a much more statistically robust and broad overview of many operations. In each case, the theme of the data pipeline remains central. While we do not introduce any advanced or new analytics algorithms or truly big data, both operational and historic data have the capacity to employ the former and become the latter.

With the widespread adoption of online digitised monitoring, smart multi-filtering systems, and the development of information technology infrastructure within the plant environment, in-house data pipeline development is well within the reach of large metallurgical processing plants. Jointly with this in-house expansion on data capabilities, rigorous control practices and management must be placed on data governance, and particularly data quality and literacy.

At smaller plants, process simulations in independent dry laboratories and scaling tests are highly vital to generate data, replicate processes and closely examine the data from metallurgical, analytics and industrial design perspectives. In addition, since legacy data predates modern data concepts, the examination of such data requires specific subject matter experts with an inside knowledge of the inner workings of the data. The value of legacy data is wholly dependent on the availability of subject matter expertise.

With the 4th industrial revolution, a proper data governance structure should coexist with rigorous data management, which embeds the value of data into master data, reference data, metadata and other various structures that are much more tolerant of expertise perturbances, including personnel changes. This highlights the immediate and pervasive need to migrate away from historic concepts of data ownership and expertise around a subject matter towards a modern implementation of data governance. Any integrated approach that intends to leverage data, and particularly multi-sourced data to optimise mineral processing should take into strong consideration the effects of a proper data governance structure and a sound data pipeline. This ensures that the integration is natural, holistic and modern, and that the data generated at each stage is cross-compatible, self-explanatory, optimally used and reused. However, in contrast to changes in technology and increases in data generation, corresponding changes in culture are much slower. Data governance, among other changes brought forth by the 4th industrial revolution requires operational, tactical, management and strategic cultural changes and should begin with a vision and strategy at the executive level. Implementation of such a strategy, including investment in cultural changes, such as provisions for training, education, employee outreach, talent recruitment and retention should coincide with technological changes. The final goal is to derive better business outcomes using technology and data, and build a company culture that promotes discovery and innovation through the adoption and utilisation of both technology and data.

In summary, we propose a holistic and transformative approach, aiming to standardise, automate and evolve the decision-making feedback loop within plants. To accomplish this, a plant would need to establish or enhance online digitised monitoring capabilities, and incorporate technologies that promote automation, such as XCT and PEPT, which are all data-laden and integrate well within the typical current plant operation. However, these techniques are often costly to acquire, maintain and operate, prohibitively so for smaller or low-profit operations, and their integration within a separate independent laboratory is probably a better approach. For all businesses of all sizes, specialised dry laboratories seem unavoidable.

With the increased data generation and analytics deployment, a well-established data bank or warehouse becomes a business necessity. Even for the independent laboratory, a proper data bank is essential. Once this approach has been adopted, all pillars of the 4th industrial revolution that depends on an abundance of quality and timely data can be realised.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.


