Theory of Sampling and QAQC enabling the application and expectations of new technology and data processing

Oscar Dominguez
Global Principal QAQC, Resources Centre of Excellence, BHP. E-mail: oscar.r.dominguez@bhp.com
Member of the World Sampling Council, The International Pierre Gy Sampling Association.

Currently, there are high expectations in the mining industry, across the Supply Chain, on how sensors and new technology providing real time data can support and optimise business decisions. In addition, sophisticated statistical algorithms, such as machine learning or conditional simulations, are more and more explored/used to address topics as uncertainty and "optimisations" in the plans, at different horizons, to "maximise the value of the business".

Despite the future of data collection is heading in the direction where sensors will be providing real time information, this is still in the development stage. The main challenge of the current status for sensors, new technology or statistical analysis, is that they are mostly based on the assumption that the data used during calibrations or data processing, is correct or representative.

This paper elaborates in more detail (with examples) on how the Theory of Sampling and the implementation of Quality Programs (QAQC & QM), across the supply chain, represent key enablers in the research, applications, selections and implementation of technology providing real time data, as well as the quality quantification of the information used as input for data processing: what a sampling protocol represents, how main and deleterious elements are distributed in the lot to be sampled, grade per grain size distribution profile, what can impact sample collection process, how gaps during sample collection shall be monitored, sources of bias, sources of variability and how this information can be used to quantify the current quality performance that will need to be improved with the technology. This paper also elaborates on the current expectations of minor/trace element data (normally on ppm levels), specifically in the understanding and challenges these types of data represent. The final objective of this analysis is to highlight the potential impacts during a capital process where new technological projects can be wrongly excluded from consideration due to errors in the baseline used for comparison, as well as the potential impact on reconciliation and marketing results due to technology or statistical analysis using biased datasets.

Introduction

Currently, there are high expectations in the mining industry, across the Supply Chain, on how sensors and new technology providing real time data can support and optimise business decisions. In addition, sophisticated statistical algorithms, such as machine learning or conditional simulations, are more and more explored/used to address topics as uncertainty and "optimisations" in the plans, at different horizons, to "maximise the value of the business".

Despite the future of data collection is heading in the direction where sensors will be providing real time information, this is still in the development stage. The main challenge of the current status for sensors, new technology or statistical analysis, is that they are mostly based on the assumption that the data used during calibrations or data processing, is correct and representative, which provides two streams of different perspective for development: 1) Have the quality of the current samples/data being quantified? Are the main sources of errors and biases being identified, understood, and considered within the baseline the new technology is going to be measured against? What is the current quality performance (bias and precision) of the process that is going to be potentially replaced by new technology? And 2) On sensors and new technology, what controls are used to measure the performance in terms of quality? Are all the sources of bias and variability considered and their impact quantified? Is the implementation of a Quality Program (QAQC) to monitor/measure the quality of the instruments considered? **(Figure 1)**

doi: https://doi.org/10.1255/tosf.130

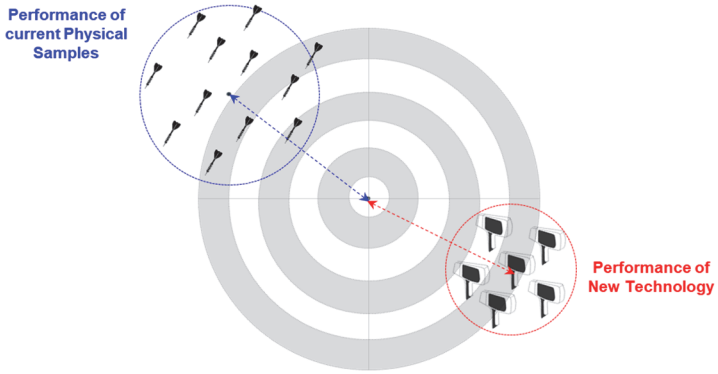

Figure 1. Scheme showing the considerations to be addressed where testing new technology versus physical samples.

Despite there is potential for a future where the collection, preparation and analysis of physical samples can be replaced by sensors and sophisticated data processing, the reality is that even in that future, the Theory of Sampling will remain relevant because it will provide the guidelines and parameters to be considered within the baseline equation to measure sensors and technology optimisations, and compare quality performance against; and also, how Quality Programs will need to be considered and adjusted to ensure sensors and new technology are delivering a representative measurement. Today, this is not always considered, and the risk for good technological tools being rejected because of subjective high expectations of the current situation, or important investment decisions on technology where maybe is not need it, is very high.

The mining industry is facing big challenges on the consistent reduction of production grades, more complex geological environment, tighter environmental requirements, and the increasing economic value and attention for minor elements at part per million (ppm) levels. In addition, we have the increasing desire in the mining industry for new technology to give us the answered we are searching for, but is the industry aware that all the sampling protocols have been optimised for the main elements require in production and minor elements or trace elements not? Are we considering the huge uncertainty related to ppm results? Is the industry clear that a ppm result is a number that an analytical methodology was able to determine, but that number is not necessarily representative to the amount found it in the ground?...

Figure 2. Sources of samples across the mining supply chain.

This paper elaborates (with some examples) on how the Theory of Sampling and the implementation of Quality Programs (QAQC & QM) across the supply chain **(Figure 2)**, represent key enablers in:

- The research, applications, selection, and implementation of technology providing real time data.
- The quality quantification of the information used as input for data processing, specifically, considerations on: what a sampling protocol represents, how main and deleterious elements are distributed in the lot to be sampled, grade per grain size distribution profile, what can impact sample collection process, how gaps during sample collection shall be monitored, sources of bias, sources of variability and how this information can be used to quantify the current quality performance that will need to be improved with the technology.
- The technical expectations for minor/trace element databases, specifically in the understanding and restrictions these types of data represent.

The final objective of this analysis is to highlight the potential impacts during a capital process where new technological projects can be wrongly excluded from consideration due to errors in the baseline used for comparison, as well as the potential impact on reconciliation and marketing results due to technology or statistical analysis using biased datasets.

Do not forget the basics: Grade per particle size distribution, the Lot DNA versus Sample DNA

The most basic concept on Theory of Sampling is that a “sample is part of a lot”, and for a sample to be called “representative”, it needs to include all the components of the lot, in the same proportions. Any deviation of this principle will generate deviations in the sample collected.

Currently, the characterisation of mining resources is performed, in general, at Head Grade level, where just a "number" is used to represent a sampling interval in a drillhole, or a production area in grade control, or an individual block in a block model, or thousands of tonnes production on a conveyor belt, or a full shipment of final product. This is where the concept of "Lot DNA and Sample DNA" becomes relevant. Lot DNA refers, on a sampling perspective, to the study and characterisation of the grade distribution (main and minor elements) across the particle size distribution **(Figure 3)**, with the objective of understanding the proportions in the lot that needs to be preserved in the sample collected (sample DNA). By having characterised the particle size distribution profile, it is going to be easier to understand the impact in the Head Grade for issues during sample collection affecting preferential particle sizes (fine or coarse fractions), especially sources of underestimation or overestimation of the main production elements, but also for minor elements and trace elements.

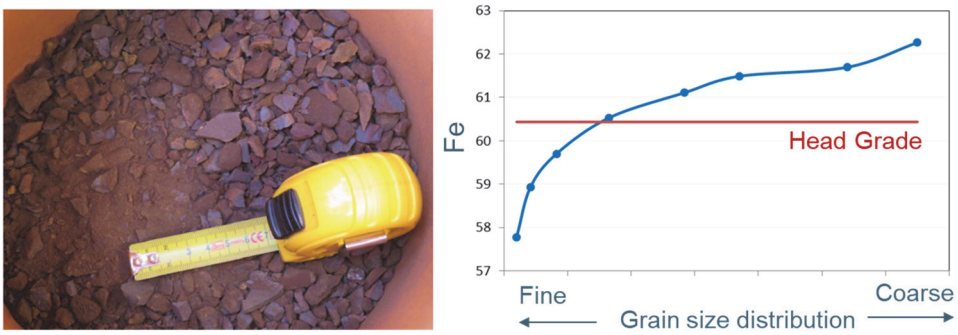

Figure 3. Example Fe grade per size distribution in an Iron Ore sample. In this example, Fe is located in the coarse part of the sample, so preferential sampling of this particle size will overestimate the Fe in the Head Grade of this sample. Note: It is also suggested to include mineralogy in the grain distribution profile, especially for geometallurgical purposes.

The importance of the understanding and consideration of the DNA will be relevant in the capacity of new technology to measure the full DNA and not just parts of it, especially when related with particular grain sizes in the lot, that will impact the representativeness (and a highly potential bias) of the measurement. Currently, it is valuable to get a real time measurement, but it is more relevant to get a representative measurement in order to optimise further decisions and final financial outcomes…

The context behind Sampling Protocols vs business expectations

In regard to Sampling Protocols, and how they can influence new technological results, it is necessary to consider and highlight the strategy established when heterogeneity studies were performed to optimise the Sampling Protocol and the Fundamental Error of the main elements required in the supply chain, where 1) in minor elements, it is expected to have a bigger Fundamental Error because they are not normally prioritised in the strategy, and 2) it is expected to have even more error in trace elements (**Figure 4**).

Example Fundamental Error (σ^2_{FE}) of Sampling Protocols on different elements in a copper mine

$$\sigma^2_{FE} = \left(\frac{1}{M_S} - \frac{1}{M_L}\right) \cdot C \cdot d^3 \approx \frac{1}{M_S} \cdot C \cdot d^3$$

UG	Protocol	CuT	Fe	As	Mo	Pb	Au	Trace Elements
1	DDH	1.6	1.9	3.3	3.6	5.6	9.5	>>10
	RC	3.2	3.6	4.3	6.1	8.9	9.9	>>>10

Figure 4. Example on how the Fundamental Error on Sampling Protocols varies per elements and their prioritisation strategy.

This is important to be highlighted, because the error and uncertainty associated with an "assay number" stored in a database will vary depending on the importance of the element for the business when the sampling protocol was designed. For example, in a copper mine, sampling protocols are expected to be optimised for copper (with a lower error), but not necessarily for minor elements or trace elements. This situation raises several points that needs to be considered when new technology is trying to be applied at minor or trace elements concentrations:

- Because the sampling protocol used to collect and prepare the samples was not designed for minor or trace elements, it needs to be highlighted that the assay result in the database it is expected to have a big uncertainty and questionable representativeness against the lot. In other words, if we have 150 ppm of a trace element, due to the sampling protocol, it is expected to have a big variability in the measurement at 95% confidence, it means, it can be150 ppm ± 50 ppm, or ±100ppm, etc, the expected standard deviation it is expected to be very important.
- Another point is related to the representativeness of these 150ppm against the original lot. Because of what has been

explained on the sampling protocol, this "assay result" has a highly questionable representativeness against the original lot. It is very unlikely that the exact 150 ppm are going to be found in the lot sampled.

- Before utilising physical samples as a reference for minor or trace elements, it is important to consider the QAQC program was used on the analytical process (ICP). Were Certified Reference Material (CRM) used and evaluated for those minor or trace elements? Has the CRM been certified by ICP?

All these points are highly suggested to be considered when new technology is trying to be applied on minor or trace elements concentrations. Results are very likely to be imprecise, but not necessarily due to issues in the sensor itself. Maybe the main source of variability is coming from the sample used in the calibration, and this variability in the absent of an understanding on the restrictions on the physical samples data, can lead to wrong decisions about the performance of the technology. This is the value of a quantified and understood quality, and restrictions of the baseline representing current situation to be potentially replaced.

Note: This is important as well on the high expectations or uses, business is trying to give to ICP data to potentially evaluate, reprocess, and recover Rare Earth, or for thresholds on ppm level defined by environmental requirements. The question is whether industry is prepared to meet the expected low variability at ppm levels?

Sources of variability and bias to be considered when trialling new technology

When trialling new technological applications across the supply chain, it is important to consider and have quantified all the potential sources of error and bias (Dominguez, 2019)[4]. Normally these sources are not considered by providers and developers, and they just rely in the number o result provided by the instrument, not quantifying or challenging the representativeness of the measurement.

Example sources of variability

Sample grain size heterogeneity is going to play a key role when trialling new sensors or technology: the coarse and heterogeneous the material to be measured is, the larger the expected variability will be **(Figure 5)**, and this is important to be considered because technological applications on conveyor belts, for example, can be dismissed because of wrong expectations of a lower variability in the results. In addition, if the full sampling process is not understood (collection, preparation, and analysis), providers can be promising better precisions against just the analytical part, but not considering the sample heterogeneity during sample collection in their equation, where ~80% of error is introduced.

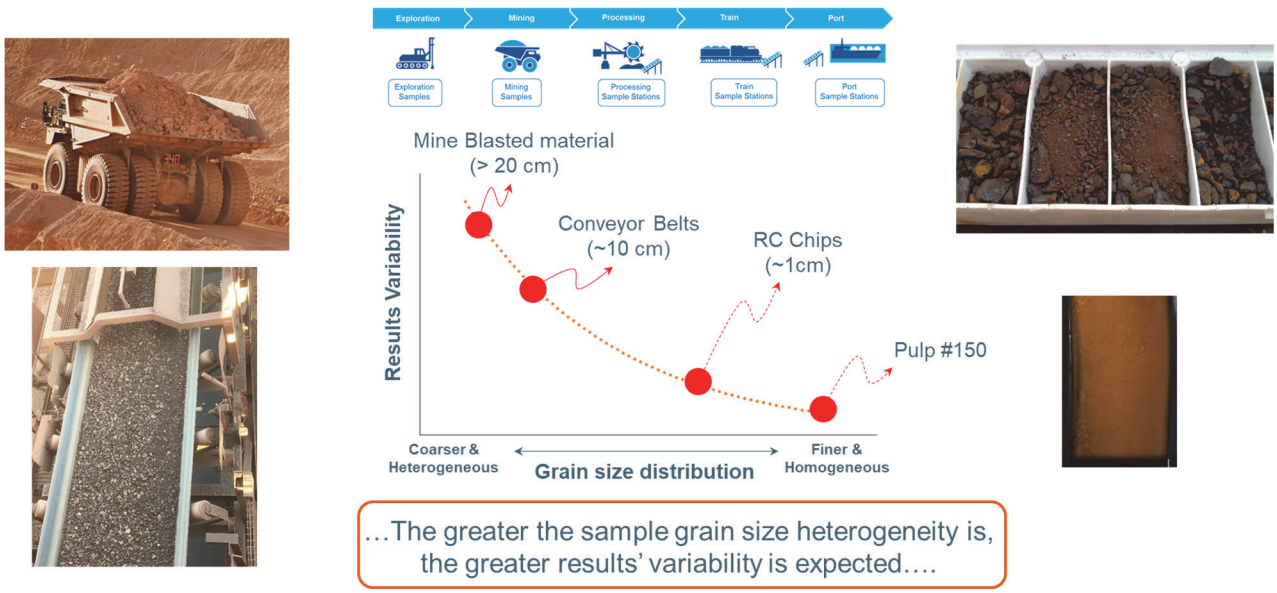

Figure 5. Example on expected variability versus sample grain size heterogeneity.

Example sources of bias

Sample segregation, when is not consider during sampling, has the potential of representing an important source of bias in the measurement of new technologies, if they are not able to cover the full stream of sample that is required to be measured. For example, in conveyor belts, if sensors performing surface measurements will only be able to cover just few microns in the sample, and if the sample is including preferential particle sizes, the measurements are going to be biased and finally will not be representative…this is another reason why it is important to have characterised the DNA of the sample, in order to know the strengths and weakness of the technology trialled.

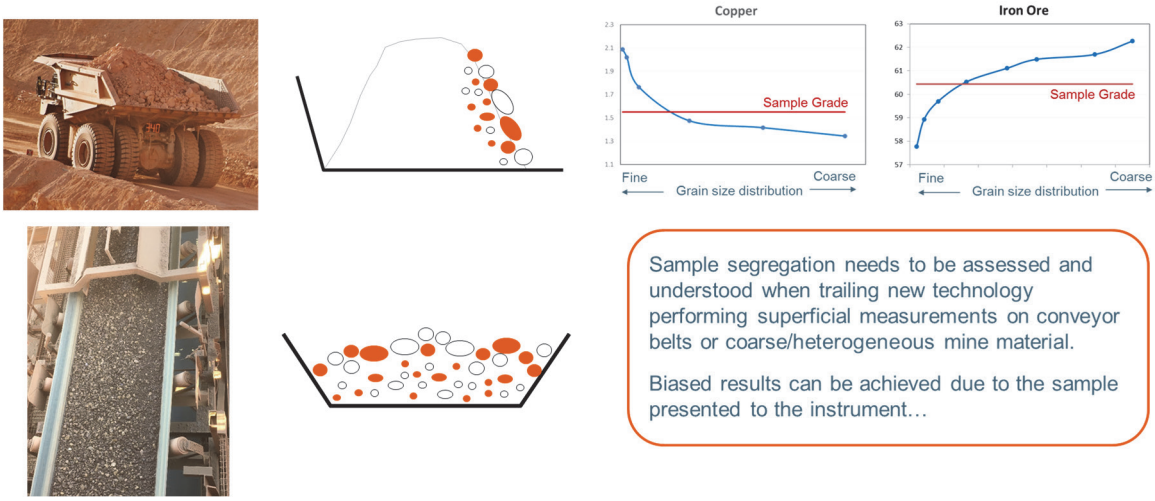

Figure 6. Example of segregation as a source of bias when new technology measurements are superficial and are applied on segregated lots.

The importance of the quality quantification of the current physical samples when is compared against new technology.

An important milestone that needs to be considered, while trialling new technological applications, is the quality quantification of the current sampling state – the physical samples – that is going to be used as a baseline comparison and are expected to be replaced by technology. It is important to have quantified the performance in terms of bias and precision (variability), because in the absence of this quantification, subjective quality expectations can lead to wrong decisions for deploying or rejecting technology.

In all the sampling points in the supply chain, it is important to have quantified the performance on sample collection, sample preparation and sample analysis, in order to have objective information to be used in the comparison and better manage potential people's subjective expectations on quality: "I want to have zero error..." and also, to have a better understanding of the inputs are going to be used for the calibration of the new technology.

This quantification request will have two potential impacts: 1) The current quality performance that will need to be beaten by new technology, and 2) The quality of the information is going to be used to calibrate the new technology.

<u>Example on sensors applied on conveyor belts</u>

As a first step, it is required to have quantified the current performance of the Sample Station: have bias and precision tests been performed? Have these parameters been quantified? Has the lot DNA been characterised? In the absent of this information, the baseline to be used as comparison, is going to be unknown and will directly impact the technical decision on proceed or not with the new technology.

In addition, in the absent of a quantified performance, technology using "dynamic calibrations", it means, technology calibrations based on the physical sample collected from the sample stations, has the increased risk of been calibrated with unknown quality data, and even worse, when factors are applied on sensors or technology to mimic samples with unknown quality.

Figure 7 shows a production report where physical samples collected from a conveyor belt and online analyser data are compared on daily basis. Report is showing a consistent difference between both sources of information. Which one is correct? In the absence of the quality quantification of the sample station, the quality of the physical samples is unknown, and also in this example, because these samples were used in "dynamic calibrations" of online analysers, the quality of this technology is also unknown. But if the quality of the physical samples has been quantified (bias and precision test performed), then is more likely an issue in the calibration of the online analyser.

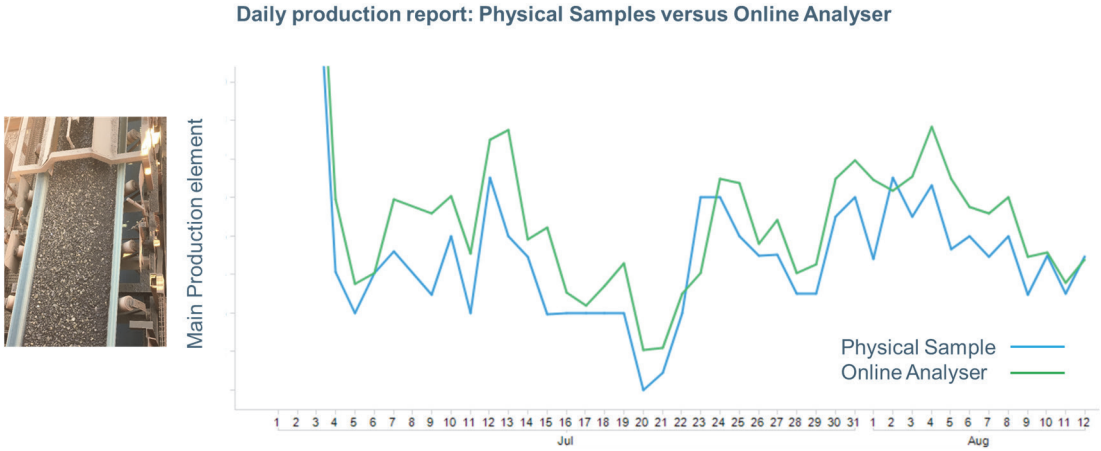

Figure 7. Daily production report showing data from Physical samples versus Online Analysers. Report is showing a consistent difference between both sources of information. Which one is correct?

Unfortunately, the current assumption in the industry is normally to say that the sensor is having calibrations issues, but this is just an assumption. What happen if the samples are biased?

In terms of the quality performance of new technology, it is highly desired to apply a “static calibration”, which refers to the use of an independent reference material to measure the performance and isolate the pure performance of the technology, as a simile of the Certified Reference Material (CRM) that are used during sample analysis in laboratories to independently measure the accuracy of the analysis.

Returning to **Figure 7**, if a static calibration has been performed in the Online Analyser, we have an independent tool supporting the performance of the sensor, so now, is more likely that the physical samples are having a negative bias. Having a more long-term view of these sensors, they represent a robust tool that can be used as a Quality Control, to monitor the information provided by physical samples.

Unfortunately, these independent measurements are not very often considered, and providers rely in the samples provided by companies and are just assumed as representative.

<u>Blasthole samples considered as reference for technology calibrations</u>

When new technology is trying to be trial on blastholes, as has been repeatable mentioned, it is very important to have quantified the current performance of the manual blasthole sampling and to have identified all the potential sources of gaps on sample collection, preparation, and analysis, in order to have robust baseline to compare new technology against.
Figure 8 shows an example in blastholes on why it is important the understanding, on a sampling perspective, of the sources of gaps impacting the quality of the manual sampling is going to be used as a baseline and inputs for calibration. In this example, a) DNA is indicating the high-grade material is located in the fine fractions, b) field inspection is showing serious gaps during sample collection, with a preferential trend towards the collection of fine fractions, c) comparison between grade control model, based on blastholes, showing a consistent overestimation of the grade against the long-term model. This information needs to be considered, because the new technology is going to be calibrated and compared against inputs with serious sampling issues.

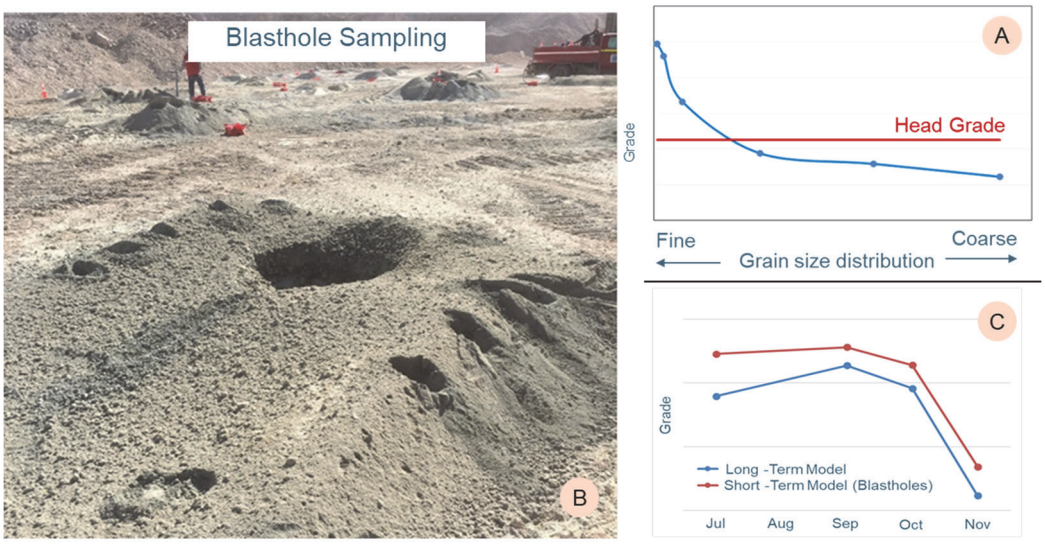

Figure 8. Example of the importance for having a good understanding of the sources of sampling gaps impacting the quality of the physical samples are going to be used as a baseline and calibration for new technology.

Similar considerations need to be taken in account when new technology is planned to be applied on drilling methodologies as Reverse Circulation (RC) or Diamond (DDH). Parameters such as Sample DNA, recovery, sample weight and the results of the QAQC program during sample collection, preparation, and analysis, that are going to allow an objective quality comparison with the new technology (Chi *et a*l, 2017)[2].

Quality Programs – QAQC – are also required on new technology

Having the context now on the big importance sampling and QAQC considerations have to define an objective baseline to compare and calibrate new technological applications, it also important now to consider what is going to be the QAQC program and Quality Management – QM – (Dominguez, 2021)[3] the new technology needs to implement, which are the restrictions of the sensors, and which are the parameters that need to be monitored.

For example, hyperspectral technology has been applied in greenfield exploration, drillholes, mine face scanning and also trailed in conveyor belt to determine the mineralogical composition of the lot. The key parameters that shall be monitored under a Quality Program are the wavelength and reflectance (Mittrup et al, 2017)[1], in order to have consistent measurements of the spectra **(Figure 9)**. Deviations in this monitoring have shown differences in the measurements between the day and the night due to temperature variations. Also, it is important to consider in the evaluation what kind of maintenance program will need to be implemented and how they are going to impact the process (in conveyor belts, for example). Finally, it is also important to know and understand the restrictions of the different methodologies: if they are a superficial measurement or if they are a volumetric measurement, because if the measurement is just few microns deep in the sample, all the theory of sampling will have an impact in the interpretation obtained.

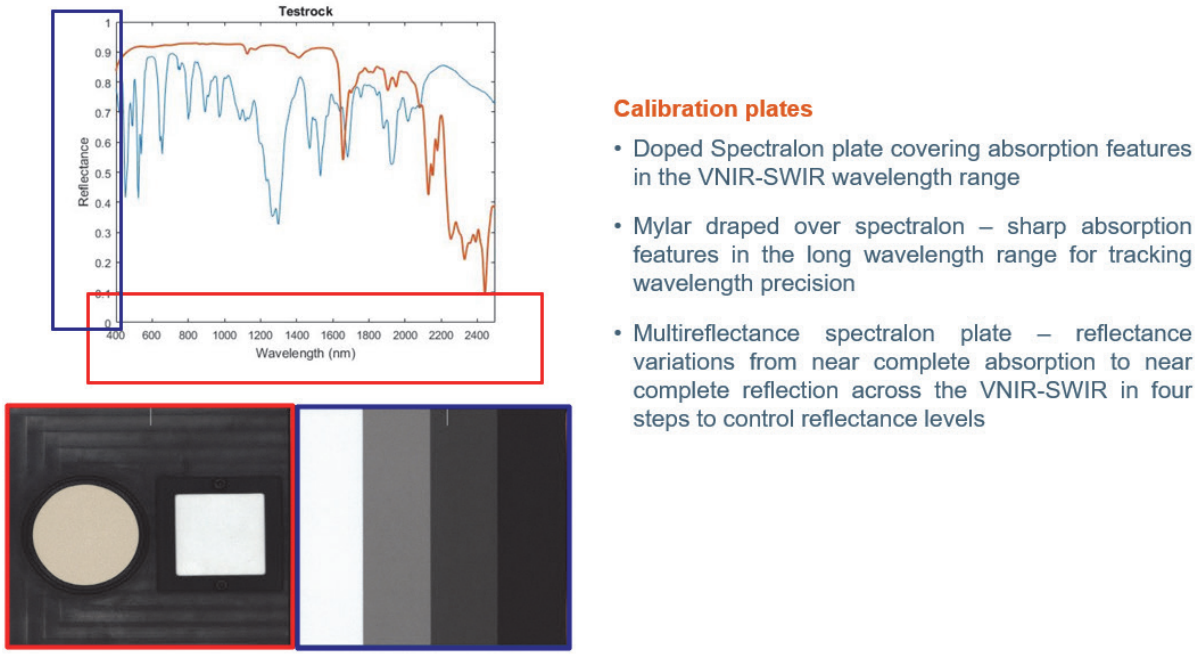

Figure 9. Example of QAQC Program applied on hyperspectral data.

Uncontrolled advanced statistics: high risk for quality

Something to be highlighted as well, in this industry desire of having real time data to perform real time decisions, is the risk of using advanced statistics, as machine learning or simulations, to mathematically close gaps in the mining supply chain through modifying factors, but no giving the business the chance to address and fix the root cause of the gaps determined in the samples/inputs affected.. This false sense of “optimisation” has the potential of hide consistent inefficiencies or gaps in the inputs (samples or technology) used in the supply chain. So, it is important to highlight the risk of using an uncontrolled advance statistic: instead, is better to have a quality quantified sampling performance based on a robust Quality Program.

Conclusions

Currently, there are high expectations in the mining industry, and across the Supply Chain, on how sensors and new technology providing real time data, can support and optimise business decisions. In addition, sophisticated statistical algorithms, such as machine learning or conditional simulations, are more and more explored/used to address topics as uncertainty and “optimisations” in the plans, at different horizons, to “maximise the value of the business”.

Despite the future of data collection is heading in the direction where sensors will be providing real time information, this paper highlights the current technical challenges to be considered in the current stage, where the main opportunity of the current status for sensors, new technology or statistical analysis, is that they are based on **just** the assumption that the samples/data used during calibrations or data processing, is correct or representative.

This paper highlighted the importance of the Theory of Sampling and Quality Programs (QAQC & QM), as an enabler for an informed technical assessment to be considered, before a new technology or data analysis is deployed, to:

- ✓ Quantify the quality of the current sampling methodology to be replaced.
- ✓ Understand the uncertainty behind the assay results stored in a database, especially for minor and trace elements.
- ✓ Support the definition of the baseline comparison between the physical sample and the new technology.
- ✓ Define a Quality Program (QAQC & QM) to be implemented in the new technology.
- ✓ Highlight risks and manage expectations for uncontrolled advanced statistical analysis.

Through some examples, this paper aimed to contribute to the understanding and suggestions on what parameters on sampling and QAQC could be considered in the assessment of new technology to technically support business desicions.

References

1. Mittrup, Dominguez & Haest (2017). A new approach to implement QAQC to technological innovations - quality on spectral data capture and process flow in the minerals industry. WCSB8.
2. Chi, B, Maddever, R A M, Mahanta, A and Dominguez, O, 2017. x000D_Near real-time assay with downhole assay tool (FastGrade 100), in Proceedings Eighth World Conference on Sampling and Blending, pp 137-144 (The Australasian Institute of Mining and Metallurgy: Melbourne).

3. Dominguez (2021). Quality Management (QM): The heart of the QAQC process. Spectroscopy World, Vol.33 No1. February 2021.
4. Dominguez (2019). Applications for Portable XRF Analyzers. PDAC 2019. Toronto, Canada.