



This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

### Le, Thi Minh Khanh; Dehaine, Quentin; Musuku, Benjamin; Schreithofer, Nóra; Dahl, Olli

## Sustainable water management in mineral processing by using multivariate variography to improve sampling procedures

Published in: Minerals Engineering

DOI: 10.1016/j.mineng.2021.107136

Published: 01/10/2021

Document Version Publisher's PDF, also known as Version of record

Published under the following license: CC BY

Please cite the original version:

Le, T. M. K., Dehaine, Q., Musuku, B., Schreithofer, N., & Dahl, O. (2021). Sustainable water management in mineral processing by using multivariate variography to improve sampling procedures. *Minerals Engineering*, *172*, Article 107136. https://doi.org/10.1016/j.mineng.2021.107136

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Contents lists available at ScienceDirect







journal homepage: www.elsevier.com/locate/mineng

# Sustainable water management in mineral processing by using multivariate variography to improve sampling procedures

Thi Minh Khanh Le<sup>a,\*</sup>, Quentin Dehaine<sup>b</sup>, Benjamin Musuku<sup>c</sup>, Nóra Schreithofer<sup>a</sup>, Olli Dahl<sup>a</sup>

<sup>a</sup> Department of Bioproducts and Biosystems, School of Chemical Engineering, Aalto University, P.O. Box 16300, 00076 Aalto, Finland

<sup>b</sup> Circular Economy Solutions Unit, Geological Survey of Finland (GTK), P.O. Box 96, FI-02151 Espoo, Finland

<sup>c</sup> Boliden Kevitsa, Kevitsantie 730, FIN-99670 Petkula, Finland

#### ARTICLE INFO

Keywords: Water quality and flotation Process water monitoring Theory of sampling Multivariogram Principal component analysis Sampling error

#### ABSTRACT

Due to the rising pressure from governmental regulations, the water recycling rate has increased significantly in mining operations over the last decades, resulting in a high variation of process water quality, which could potentially impact the plant performance. The current effort to assess water quality in mining is shifting from managing water to fulfill environmental regulations (focus on the effluents) to controlling water quality to maintain the operating performance (focus on the water within the process). However, minimal effort has been made to design a dedicated sampling procedure for process water. This study investigates the use of multivariate variography and principal component analysis (PCA) for improving the process water sampling procedure at the Kevitsa Mine, Finland. The aim is to design a sampling procedure for evaluating water quality using two different types of datasets and illustrating the impact of the dataset structure on the sampling design. The results showed that the common spot sampling procedure generated a very high sampling error and was not the best practice for process water. The weekly sampling frequency used at the mine site, suitable for fulfilling environmental regulations was too low to capture the process water variation. Therefore, it is not recommended to use environmental water datasets for operating control purposes. The multivariate variographic analysis revealed the hidden cyclic variation through its ability to summarize the time variations and the correlation between multiple variables that were not visible through the classical univariate variogram approach. However, the number of increments recommended by the global multivariogram became impractically high. Hence, an alternative approach combining PCA to the mutivariogram was used to filter noise from the data and keep the relevant information. This study highlights the benefits of using multivariate variography to improve water sampling procedures in the mining industry and to reduce both operational and environmental risks associated with water quality variability. Thus, this method has the potential to be used in worldwide mining operations as a standard procedure for sampling water to provide reliable results.

#### 1. Introduction

Water quality assessment is a critical aspect of mining operations in fulfilling environmental regulations and controlling plant performance (Mudd, 2010; Kinnunen et al., 2021). It is clear that for an environmental approach, a suitable monitoring program is essential for reducing the risks of pollution related to water (Bezuidenhout, 2009; Muniruzzaman et al., 2018). For such objectives, the data should be collected regularly in order to identify long- term trends, periodic changes, and fluctuations in rates of changes (Canada Environment, 2009). In terms of process control, water quality plays an essential role

on flotation performance and needs to be monitored carefully, especially when process water is reused or recycled (Forssberg and Hallin, 1988; Levay et al., 2001; Johnson, 2003; Liu et al., 2013). Nevertheless, efforts to build water sampling procedures for process performance control remain limited. When performed for process performance purposes, the sampling frequency needs to be high enough to capture the variability of process water properties that need to be monitored. The latter are defined according to their relevance to the process and are often case-specific.

The sampling errors are significant sources of uncertainty when reporting a measurement (Holmes, 2004). In the case of process water,

https://doi.org/10.1016/j.mineng.2021.107136

Received 14 February 2021; Received in revised form 25 May 2021; Accepted 9 August 2021 Available online 27 August 2021

0892-6875/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

<sup>\*</sup> Corresponding author. *E-mail address:* thi.5.le@aalto.fi (T.M.K. Le).



**Fig. 1.** Typical positively increasing variogram. The experimental nugget effect  $(V_0)$  is the value of the variogram at the origin, the range is the lag distance beyond which no further autocorrelation occurs, and the sill which reflects the overall process variance is the flat part of the variogram.

collecting a representative sample is a prerequisite to ensure a reliable result. The most commonly used method is spot (bottle) sampling (Madrid and Zayas, 2007), consisting of collecting a single sample at a given time and location in the process. The obtained water sample is then stabilized and sent for analysis under appropriate preservation conditions, i.e., darkness and low temperature. The information thus obtained is unique to the sampled process stream at the time of sampling. Therefore, it seems obvious that such an approach could not give a representative picture of the high variation of water quality over time (Schumann et al., 2009; Le et al., 2020a).

The Theory of Sampling (TOS) provides a framework for good sampling practice and a set of rules that aim to eliminate sampling biases and minimize sampling errors (Gy, 1971, 1992, 1998; Pitard, 1993). One of the most important developments of the TOS for the mining industry is applying the semi-variogram (denoted as variogram in this paper) to characterize the autocorrelation and the heterogeneity of process streams and to quantify the sampling errors (Saunders et al., 1989; Minkkinen, 2004). Process water streams moving through a pipe could be considered one-dimensional (1D) objects. The most appropriate method to capture the distinct spatial or temporal correlation of 1D object is composite or aggregate sampling. Small increments consisting of a full cross-section of the stream are collected for a fraction of time at several time intervals during the sampling period and then aggregated to make the final sample (Esbensen et al., 2007). Collection of increments can be performed following three different sampling modes, i.e., random (R), stratified (SR), or systematic (S) sampling (Minkkinen, 2004). The choice of the sampling mode is crucial because it changes the variance of the lot's mean. If the same number of samples is extracted from the lot, systematic sampling usually gives the lowest variance of the lot mean (Saunders et al., 1989). Sampling program designs for variographic data modeling use, as a standard practice, the systematic mode and leave each increment as a discrete sample.

The shape of the variogram provides numerous information on the complexity of the sampled lots variation via three main features: the range, the sill, and the nugget effects (Fig. 1). The range indicates the limit of the inter-lag distance after which autocorrelation is no longer observable. The sill informs on the total variability of the lot for the properties considered. Furthermore, the nugget effect estimates the total variance not related to the process variability but only to random effects, i.e. sampling, preparation, and analytical errors. The shape of the variogram reflects the characteristics of the process. Variograms of processes displaying a drift are mostly increasing variogram, which may or may not level off after the range (Minkkinen and Esbensen, 2014).

The variographic approach used in the mining and mineral processing industries has always been limited to one variable at a time (Esbensen, 2017). Indeed, even when samples must be representative for

Table 1							
<b>Dverview o</b>	f the two process water c	latasets used in this s	tudy.				
Dataset	Sampling location	Period	Duration	Measurement frequency (j)	Number of observations (N)	Number of variables (p)	Variables/ Properties (X)
1 - HF	Main process water	7-14/3/2018	1 week	10 min	980	5	SPC, pH, ORP, turbidity, DO
	stream						
2 -	Output of the tailings	29/5/2017-30/7/	14	1 week	70	12	pH, SPC, sulfate, thiosulfate, nitrogen, phosphorus, calcium, magnesium,
Weekly	puod	2018	months				manganese, nickel, iron, silica

a range of properties, the standard solution is to reduce the problem to a single variogram by identifying the property with the highest heterogeneity contribution and designing the sampling procedure on this property only. This approach is known to underestimate the global sampling variance by not considering the multivariate aspect of heterogeneity and correlations between individual properties (Dehaine et al., 2016). In the case there are multiple properties of interest, the use of the multivariate variogram (denoted as multivariogram) for sampling design is recommended. The application of multivariate variography in TOS for temporal data analysis and sampling design was firstly introduced by Dehaine and Filippov (2015). This study showed that the multivariogram could complement the univariate variogram by summarizing global time variation for multiple properties of interest. However, the resulting multivariate global variance may be relatively high and lead to a high number of increments to be sampled to make the final representative sample of the lot. Minkkinen and Esbensen (2014) have shown that computing the variogram on the first few principal components of a Principal Component Analysis (PCA) allows combining multivariate and variographic analyses in a way that explains variables covariance and the distinct spatial patterns of 1D lots. Therefore, another alternative was suggested, combining PCA with the multivariogram by applying the latter to the PCA scores instead of the raw data (Dehaine et al., 2016). As the PCA reduces the dimensionality of the data and filters the noise (random effects), this method reduces the global variance while summarizing the essential information about variability.

This paper investigates the application of multivariogram and variographic analysis to design a suitable sampling procedure for water quality monitoring and assessment. To do so, two datasets, collected from the same active mining operation and representing two distinct time scales and applications, e.g., a high-frequency process control dataset and a weekly environmental monitoring dataset, have been used to illustrate the impact of the initial dataset structure on the sampling design.

#### 2. Materials and methods

#### 2.1. Datasets

The data used in this study was obtained from the Boliden Kevitsa Mine located in Northern Finland. The plant processes a Ni-Cu(-Co) sulfide ore by froth flotation. The mine water handling consists of a water recycling circuit recirculating more than 90% of the tailings dam water back to the plant.

Two datasets were provided: a high-frequency (HF) dataset and a weekly monitoring dataset (Table 1). It is important to recall that the designed sampling procedure will always have a lower frequency than the dataset used for the design. For example, the HF dataset (one reading every 10 min) could be used to design daily sampling procedures. Meanwhile, the weekly sampling dataset is used to design the monthly sampling procedures, suitable for environmental monitoring purposes only.

Traditionally, water pollution regulation has predominantly been focusing on pH, toxic metals, metalloids, sulfate (Opitz and Timms, 2016), hardness as given by the calcium and magnesium concentrations (Fowlie et al., 2001), nutrients (phosphorus and nitrogen), and lately thiosalts due to their capacity of acidifying and reducing dissolved oxygen of the receiving water body (Miranda-Trevino et al., 2013). Therefore, the weekly dataset used to design sampling procedures for the environmental purpose only considers these properties. The HF dataset included all properties that could be monitored at high frequency by implementing the YSI ProDSS multiprobe (Xylem Inc., USA) in the water process. In total five properties were monitored: Specific Conductance (SPC), pH, Oxidation-Reduction Potential (ORP), turbidity, Dissolved Oxygen (DO).

Both datasets were processed before conducting the multivariate

analyses. First, each variable was auto-scaled to obtain dimensionless values as the results from variographic and PCA analyses depended on scaling. Next, as outliers have detrimental effects on variographic analyses (Esbensen et al., 2007), two methods were used to detect and remove outliers from the dataset. The reason for using two different methods was the potential seasonal variation of the weekly dataset. As long-range variations were expected for the weekly dataset, the method for removing outliers in short-range variation was not applicable. For the HF dataset, observations outside the range of the mean  $\pm 2\sigma$  threshold were considered as outliers while for the weekly dataset, observations with high value of Q-statistics and Hotelling's T<sup>2</sup>-statistics (see Section 2.3) from the PCA model were considered as outliers. The threshold of Q-statistics and Hotelling's T<sup>2</sup>-statistics to define outliers was set at 95% confidence limits.

The HF dataset contained a total of 980 observations. Appendix Table A1 presents the summary statistics of all properties measured in the HF dataset. The weekly monitoring dataset contained 60 observations. The water samples were taken from the tailing ponds discharge stream, which feeds the process water tank supplying the plant, using spot sampling. Appendix Table A2 presents the summary statistics of all properties measured in the weekly dataset.

#### 2.2. Multivariate variographic analysis

The general theoretical methodology for variographic characterization, and the relative variogram calculation and sampling errors, have been fully described in the literature (Pitard, 1993; Gy, 1998, 2004; Petersen and Esbensen, 2005) and will therefore not be described here. In multivariate variography, each variable measured is considered together in one multivariate dataset and combined in one vector, X, which constitutes of the p individual variables. The multivariogram Vj of X is then calculated using the Eq. (1) (Dehaine et al., 2016):

$$V_{j}(X) = \frac{1}{2(N-j)} \sum_{i}^{N-j} (X_{i} - X_{i+j}) M (X_{i} - X_{i+j})^{T}, j = 1, \cdots, N/2$$
(1)

where the subscript <sup>T</sup> is the transpose operator, N the total number of increments (i.e., samples) collected, j the lag parameter, and M a metric (positive definite  $p \times p$  matrix) which defines the "distances" between the units i.e., the relationship between the variables. The Mahalanobis distance (MD), for which M is defined as the inverse of the variance-covariance matrix of X, has been used as it takes into account the correlation in the data (Bourgault and Marcotte, 1991; De Maesschalck et al., 2000) and is considered to be adapted to multivariate variography (Dehaine and Filippov, 2015; Dehaine et al., 2016). All (multi-)variograms were calculated using the same algorithm as in Dehaine et al. (2016) based on the tutorial by Petersen and Esbensen (2005) for univariate variograms. The algorithm took raw data as input and transformed them into heterogeneity contribution (i.e the matrix X) for (multi-) variograms calculation purposes. The definition of heterogeneity contribution have been fully described in the literature (Petersen and Esbensen, 2005; Esbensen et al., 2007). The nugget effect (V<sub>0</sub>) was estimated by backward extrapolation of the variogram towards the origin.

#### 2.3. Principal component analysis (PCA)

PCA is a well-known variable reduction and data compression method in chemometrics (Wold et al., 1987; Wise & Gallagher, 1996; Bro and Smilde, 2014). PCA aims to simplify the dataset by transforming the original variables or axes into new uncorrelated principal components (PCs). Each PC is a linear combination of the original variables (Wold et al., 1987; Omo-Irabor et al., 2008) which means it describes potential correlations within the original variables.

PCA consists of decomposing the initial matrix data to the score, loadings, and residual matrixes. The score matrix contains information



Fig. 2. Variation over time of the 5 properties of the HF dataset (auto-scaled data).

on how different observations relate to each other while the loadings matrix contains information about how variables relate to each other. Two diagnostic tools could be calculated from the PCA model, to identify potential outliers: the Q-statistics and Hotelling's T<sup>2</sup>-statistics (Wise and Gallagher, 1996). The multivariate data analysis in this study was performed with Matlab® (version R2017b, MathWorks Inc.) and PLS toolbox.

#### 2.4. Nyquist frequency for sampling

The Nyquist frequency is the minimum frequency of sampling required if the variable of interest possesses a cyclic variation with a given frequency. In theory, this frequency is defined as twice as the frequency of the cycle of interest. In practice, it is preferable to have a sampling frequency of 5–10 times the frequency of the cycle of interest (Napier-Munn, 2014). For example, if the cycle of variation is ten weeks, the minimum sampling frequency should be about one every 1–2 weeks.

#### 2.5. Methodology

Two variographic approaches were used for designing the sampling procedure for the mine of interest: The univariate and the multivariate approach. The univariate approach consisted of applying variographic analysis on the individual properties while the multivariate approach consisted of applying the multivariate variographic analysis on multiple properties simultaneously or on the first few principal components that summarized the variation of the whole dataset. Different approaches would provide different possible sampling procedures. However, based on the objective of the sampling, the practical constraints, the relationship between variables, and relations between PCs, a final sampling procedure would be selected.

#### 3. Results

#### 3.1. HF data for process water control

#### 3.1.1. Raw data analysis

The variation over-time of the five properties of interest of the HF dataset is illustrated via the auto-scaled data in Fig. 2. Similar trends were observed for (1) SPC and pH, and (2) ORP and DO. The correlation between pH and SPC will be explained in detail in Section 3.1.3. The correlation between ORP and DO was expected as oxygen was the strongest oxidant in the media, so DO variation would directly affect the redox condition and hence the ORP. Turbidity had a completely different behavior than the other four properties with a wide range of fluctuations and no noticeable sudden changes as other properties.

#### 3.1.2. Univariate variographic analysis

The individual variograms for the five properties of the HF dataset are presented in Fig. 3.



Fig. 3. Univariate variograms of the five properties of the HF dataset. Two main groups of variograms were observed: a high-sill variables group (turbidity and ORP) and low-sill variables group (SPC, pH, DO).



Fig. 4. Scores (left), variable loadings (middle), and score variograms (right) of the first two PCs explaining more than 95% of the variability of the HF dataset.

Two main groups of variograms were observed: a high-sill variables group (turbidity and ORP) and low-sill variables (SPC, pH, DO). The variograms of pH, ORP, and DO showed a typical increasing variogram shape while the variogram of SPC displayed cyclic behavior with a period of j = 380 = 3800 min = 2.6 days. On another hand, the variogram of turbidity showed a large scale increasing trend over the monitoring period. This trend continued out of our study range. In this case, de-trending was needed to remove the trend and therefore show possible cyclicities. The variogram of the detrended turbidity is found in Appendix Fig. A3. It showed clearly a typical cyclic variation with the periodicity of j = 120-140 = 20-23 hrs. This frequency was clearly associated with the daily change of tailings discharge location in the pond. In addition, the turbidity de-trended variogram showed a range of approximately j = 100 = 1000 min = 16.7 hrs while a common range of approximately 300-350 lags (2-2.5 days) was observed for SPC, pH, ORP, and DO. The shorter range observed for turbidity, suggesting that turbidity might require a higher frequency of sampling.

Because ORP was the property displaying the highest variability, the common procedure for the univariate variographic approach was to design the sampling protocol based on this property only. The corresponding standard deviation of sampling error (SDSE) for ORP was calculated to get information on the sampling strategy and the number of increments that should be sampled (Appendix Fig. A1). In terms of sampling mode, systematic sampling gave the lowest variance. Two or three increments were already enough to reduce the SDSE to below 10% (Appendix Fig. A1). The ORP variogram range suggested that samples must be collected at least every 2.5 days systematically, or more often if more resolved insights into the process variation are deemed of interest.

This sampling protocol, while being optimum to monitor ORP, was certainly not optimum for the other properties. Indeed, this univariate analysis deliberately ignored the other properties, particularly turbidity, the variogram of which has a shorter range and, along with the one of SPC, shows signs of potential cyclic variations. Therefore, designing the process water sampling campaign solely on ORP will ultimately result in underestimating the overall global sampling variance for all the other properties that may not contribute as much as ORP to the latter but which may well be more important in terms of process control. Thus, in such cases, the use of multivariate approaches such as PCA and multivariograms must be applied (Dehaine et al., 2016).

#### 3.1.3. Multivariate variographic analysis

PCA was used to summarize the overall variability of the dataset in a

reduced number of components, to identify trends in the variations of the studied physico-chemical properties and their relations. The two first PCs, which explained more than 95% of the overall variability (77% and 18%, respectively), were chosen to build the model (Fig. A2).

Fig. 4 shows the PCA loadings and scores plot along with the two first PCs. PC1 mostly captured the changes in process water quality around observation 500. The loading plot indicated that PC1 accounts for the variability of all the properties (in particular pH, ORP and DO). Precisely, ORP and DO display positive loadings, which was expected based on the previously observed correlation between these variables, while SPC, pH, and, turbidity display negative loadings on PC1. The negative correlation between ORP and pH was in agreement with the Pourbaix diagram of water indicating that in an equilibrium state of water, pH, and ORP had a negative correlation (Greet, 2010). The positive correlation between SPC, pH, and turbidity was frequently observed during wintertime for this specific mine site. During that time of the year, the ice cap blocked the majority of the tailings pond, the residence time of the water in the pond could oscillate between long and short in several days. When the residence time was short (due to the change in discharge point or shortcut channel created in the pond), the tailings water which was high in pH, suspended solids, and dissolved solids returned quickly to the processing plant. Therefore, those parameters often varied together and were positively correlated. The score value on the PC1 experienced a sudden jump from a negative value to a positive value around observation 500. Then it decreased slowly to a near-zero negative value around observation 800. Some events had impacted the process-water quality after observation 500. This event was mostly characterized by an increase of ORP, DO, and a decrease of SPC, pH, and turbidity. PC2 mainly describes variation in SPC and turbidity. The loading plot showed that turbidity had positive loadings while SPC had negative loadings on PC2. At least three cycles of variations can be observed in the PC2 score plot, from observation 0 to observation 600. After that, the scores tended to decrease. The negative correlation between turbidity and SPC was unexpected and the reason behind remained unclear. However, the most likely reason was the precipitation due to sudden changes in thermodynamic conditions, which decreased the dissolved matter in the stream and at the same time increased the turbidity.

The PCA scores can be used to perform the variographic analysis in fewer variograms instead of calculating a variogram separately for each property as done before. It benefits from highlighting hidden trends in the data which are not visible with the individual variograms. The



Fig. 5. Multivariogram applied to the whole HF dataset (top left) and to the PC1-2 scores (bottom left). The corresponding sampling error associated with the three sampling modes (R: Random, SR: Stratified random, S: Systematic) is expressed as the standard deviation of the sampling error (SDSE, middle) and decrease of SDSE (left) as a function of the number of increments collected (Nu) for the HF data.

variograms of the individual PC1 and PC2 scores are shown in Fig. 4. The comparison between the scores plot and the corresponding variogram suggested that the variograms filtered out the random variance (the short-range variations visible in Fig. 4) while modeling mainly the longrange phenomena (drift and periodic events). Both variograms showed increasing trends, indicating that their scores differed from random (i.e., flat variogram). The PC1 scores variogram reflected the global trend for all physico-chemical properties in the process (the non-random continuous part of variation). The PC2 scores variogram described the longrange periodic fluctuation trend (the non-random, continuous, cyclic part of the variation) of SPC and turbidity. Three local minimums were observed at around j = 220-230, 420-440, and 580-600. Such observations suggested a potential cyclic phenomenon with a period of approximately 220–270 lags = 2200–2500 min  $\approx$  1.5–1.7 days. There were no other obvious changes in operational practices at the mine site that can explain a cyclic phenomenon for such a short period than the effect of residence time in tailing ponds.

The multivariograms applied to the whole HF dataset, referred to as global multivariogram in Dehaine et al. (2016), and the scores of the two first PCs are shown in Fig. 5 along with the corresponding SDSE profile as a function of the sampling mode and the number of increments collected.

Both multivariograms displayed a higher sill than any of the univariate variograms in Fig. 3. It is a known consequence of considering multiple properties for the variographic analysis with metrics taking into account correlations between the variables (Dehaine et al., 2016). The global multivariogram displayed a classical increasing variogram shape with a range around j = 250 = 1.7 days. In winter, the global composition of the process-water may change every 1.7 days, which may relate to the residence time of water in the tailings pond. This residence time is shorter in winter since the pond's available volume gets reduced by the formation of the ice cap. Fig. 5 shows that the SDSE associated with the spot sample (i.e., only one increment collected) is relatively high. An aggregate sample with only two increments could reduce the SDSE by almost 60%, and with four increments, the SDSE was decreased by 80%. However, to reduce the SDSE to below 10%, at least 15 increments must be taken to make the final sample, which was relatively high and difficult to implement in practice. Additionally, the global multivariogram (upper-right) did not clearly capture cyclic variation observed on the SPC and turbidity individual variogram and the PC2 scores variogram.

Applying the multivariogram to PC1 and PC2 scores (referred to as a multivariogram of PC1-2 in the text) decreased the multivariogram sill for the HF data by half (Fig. 5, bottom). Additionally, the nugget effect, which was quite significant in the global multivariogram, was reduced to near-zero with the multivariogram of PC1-2. In TOS, the nugget effect reflects the short-scale process variability and includes all of the sampling, analytical and sample preparation variances (François-Bongarcon, 2004). This result must be related to the nature of the dataset which only contains the most significant and important variance of the original dataset since PCA acted as a filter on the raw data to retain only relevant information on the global variability of the dataset. Hence, the multivariogram applied to the first PCs provides a more representative, i.e., noiseless, summary of the structure of variability than multivariogram applied to raw data (Dehaine et al., 2016). In a TOS point of view, the assumption of using only few PCs to modeling is that they represented the variation that purely attributed to 1D process variation (i.e Time fluctuation error and cyclic fluctuation error). The remaining variance from the PCs that are not taken into consideration is considered as related to 0D variation (f.ex. incorrect sampling error, total analytical error, etc.) The addition of those "noise" elements in the modeling will only increase the nugget but did not affect the general shape of the multivariogram.

Another interesting point was that the PCs scores multivariogram captured the cyclic phenomenon observed previously in the variogram of PC2 and that can be attributed to SPC and turbidity. According to the range of multivariogram of PC1-2, at least one sample must be collected during a period of j = 380 = 3800 min = 2.6 days. As suggested by the SDSE profile, this sample must be aggregated from five increments to decrease the standard deviation below 10% (Fig. 5, bottom).

#### 3.2. Weekly sampling data for environmental monitoring

#### 3.2.1. Raw data analysis

Fig. 6 shows the variation over-time of the 12 properties of interest of the weekly dataset. A tendency of seasonal variation was observed for most properties, particularly for sulfate, thiosulfate, pH, and metal contents. High short-scale variability was observed for nitrogen, phosphorus, calcium, and magnesium. These four properties were all greatly influenced by the flotation process in the plant since most of the



**Fig. 6.** Variation over time of the 12 properties of the weekly dataset (autoscaled data) with the corresponding temperature profile (background). (A) Reagents related elements. (B) Anions and SPC. (C) Metals, metalloid, and pH.

phosphorus, calcium, and magnesium content of the process-water was coming from the reagent used during flotation. Nitrogen is a known residue from ore blasting.

#### 3.2.2. Variogram on raw data

The individual variograms for all the properties of the weekly dataset are presented in Fig. 7. In practice, it is not recommended to compute the variogram further than j = N/2 (Esbensen et al., 2007). However, some of the variograms displayed a potential cyclic fluctuation over j =35 = N/2. Therefore, to capture the full cycle, the authors chose to display the variogram until N-j = 20 (i.e., 50 lags). The cyclic frequencies (long and short) of each element/compound are summarized in Table 2.

Results indicated a high-sill variables group composed of thiosulfate, nitrogen, manganese, nickel, and iron. Variograms for thiosulfate, nitrogen, magnesium, manganese, nickel, and silica displayed a very similar tendency, characterized by a long-range cyclic variogram shape with a period around j = 45 lags = 11 months. Among them, magnesium had an additional shorter-range cyclic variation with a period of j = 10 lags = 2.3 months.

Variograms for SPC, pH, sulfate, phosphorus, and calcium suggested a common cyclic phenomenon with a shorter-range period around j = 15–25 weeks  $\approx$  4–6 months. Since calcium and phosphorus levels in process-water were directly linked to the reagents used in the flotation plant, this medium-range cyclic fluctuation could be a consequence of adapting the operating conditions in the plant with the seasonal variations and the storage time in the main tailings pond. Additionally, phosphorus and calcium possessed an additional short-range cyclic variation characterized by a j = 2.5 = 17.5 days. Those variations might reflect the adjustments of operating conditions due to ore variation.

The variogram for iron showed only a short period cyclic variation of  $j=4-5\approx 1$  month. The Kevitsa mine feeds the processing plant on an ore-block by ore-block basis, each block taking a week or more to be processed. Some ore is temporarily stored (over one month) at the ROM (Run Of Mine) pad. This observed cyclic variation may have been due to changes in ore mineralogy and varying state of oxidation of ore coming from the ROM pad and storage stockpiles.

As total iron assays displayed the highest variability, the recommended sampling protocol for iron was to use stratified random sampling or systematic sampling with at least 5 increments (to reduce the SDSE to below 10%) and a sampling frequency at a rate of one sample per week due to its short cyclic behavior.

#### 3.2.3. Multivariate variographic analysis

The three first PCs were chosen to build the model as they explained more than 81% of the variation of the dataset (44%, 30%, and 7%, respectively) while the remaining PCs did not provide any meaningful interpretation of the data (Fig. 8).

The 3D scores plot (Fig. 8) showed a tendency of clustering between summer and winter observation (i.e., from June to September vs the rest of the year). In summer, the concentration of metals tended to increase in the process-water while the SPC and concentration other compounds concentration tended to decrease. An opposite trend was observed for winter, characterized by an increase in pH, SPC, and decreased the metal concentration. The 3D loadings plot indicated a negative correlation between metals, metalloids, and pH. This correlation was consistent with the thermodynamic relation between metal-hydroxide solubility and pH (Boardman et al., 2004; Kuyucak, 2006). The correlations between SPC, calcium, and sulfate were expected since calcium and sulfate were among the major compounds found in the process-water. A positive correlation between magnesium, nitrogen, and thiosulfate was observed. However, the reason underlying this correlation remained unclear.

Fig. 9 shows the scores and loadings plot for the three first PCs.

As suggested by the loadings, PC1 mostly captured the changes in thiosulfate, nitrogen, metals, and metalloids concentrations. Precisely, nitrogen and thiosulfate displayed strong positive loadings while metals and metalloids displayed negative loadings for PC1. The scores of PC1 showed a gradual decrease during the summer-autumn time and increased again during winter-spring. It suggested that, during the summer-autumn period, the concentration of nitrogen and pH in the process-water decreased gradually and increased again during the winter-spring period. The opposite tendency was observed for variations in metals and metalloids concentration. The latter was explained by higher temperatures favoring the oxidation of the ore and the activities of bacteria resulting in higher levels of dissolved metals and metalloids in the process-water. The variable profile (Fig. 6) confirmed the thiosulfate concentration decrease over the summer-autumn period and increased again during the winter-spring time. This result may seem counter-intuitive as the literature indicates that the oxidation of sulfide to thiosalts is enhanced by temperature (Miranda-Trevino et al., 2013). However, it is important to note that since the speciation of other thiosalts was not analyzed, it was impossible to conclude that the overall thiosalt concentration decreased during summertime. Additionally, as thiosalts speciation strongly depends on pH, the pH variations can also affect thiosulfate concentration (Miranda-Trevino et al., 2013). Besides that, thiosulfate is not the only thiosalt compounds present in the mine waters, and it is even not the major one (Whaley-Martin et al., 2020).

PC2 mainly described variation in sulfate, phosphorus, calcium, magnesium, and SPC, all positively loading PC2. The score plot



Fig. 7. Univariate variograms of the 12 properties monitored in the weekly data.

## Table 2 Difference cycle of variation observed in the univariate variogram of the 12 properties monitored in the weekly data. The sampling frequency is 5 times the frequency of the cycle of interest.

		SO4 <sup>2-</sup>	S <sub>2</sub> O <sub>3</sub> <sup>2-</sup>	Ν	Р	Ca	Mg	Mn	Ni	Si	Fe	pН	SPC
Long cycle frequency	Lag	23	40	45	15	25	45	45	45	45	-	25	24
	Month	5	9	11	4	6	11	11	11	11	-	6	5
	Sampling Frequency (weeks)	5	8	9	3	5	9	9	9	9		5	5
Short cycle frequency	Lag	-	-	-	2.5	2.5	10	-	-	-	4.5	-	-
	Month	-	-	-	0.6	0.6	2.3	-	-	-	1.1	-	-
	Sampling Frequency (Days)	-	-	-	3.5	3.5	14	-	-	-	6.3	-	-



Fig. 8. PCA analysis of the weekly dataset. Eigenvalues of the principal components (left), 3D plot of loadings (middle), and scores (right) projected on the PC1 vs PC2 vs PC3 component plane.



Fig. 9. Scores with background temperature profile (left) and variable loadings (middle) and score variogram of the first three PCs explaining more than 81% of the data variability of the weekly dataset.



Fig. 10. Multivariogram applied to the whole weekly dataset (top left) and to the PC1-3 scores (bottom left). The corresponding sampling error is expressed as the standard deviation of the sampling error (SDSE, middle) and decrease of SDSE (left) as a function of the number of increments collected (Nu) for the weekly data.

indicated a slow increase in the value of the scores from May to March 2017, followed by a sharp decrease from March to May 2018, and an increase again (Fig. 9). This result was believed to be the direct

consequence of the formation of an ice cap at the top surface of the tailings ponds surface, reducing the availability of water in the tailings pond. During the summer, the ice melted down and diluted the tailing



Fig. 11. PCA analysis of the univariate variogram. 2D plot of loadings projected on the PC1 vs PC2 factorial plane (left). Scores (middle), and variable loadings (right) of the first two PCs explain more than 85% of the data variability of the univariate variogram.



Fig. 12. Multivariogram applied to different group of parameters: Iron, Phosphorus, Calcium (top left), SPC, Sulphate, Nickel (middle left), other parameters (bottom left). The corresponding sampling error is expressed as the standard deviation of the sampling error (SDSE, middle) and decrease of SDSE (left) as a function of the number of increments collected (Nu) for the weekly data.

water again. The positive correlation between sulfate, calcium, magnesium, and SPC was expected since they are the major compounds/ elements of the water matrix. It is important to note that even if both PC1 and PC2 described a seasonal variation, their phases of variation were not identical. The local minimums were observed around September for the PC1 while they were in May for the PC2. Therefore, the mechanisms that were responsible for those cyclic variations were not similar.

PC3 was mostly positively loaded by iron and pH and thus represented the variations of these properties. The PC3 scores plot was noisier than the PC1 and PC2 score plots, but a short-period cyclic tendency can still be observed (Fig. 9).

All the variograms, the PC1, PC2, and PC3 showed cyclic behaviors with distinct characteristic periods. All scores variograms reflected the long-range (yearly) periodic fluctuation. However, the PC2 variogram showed an additional medium-range (j =  $20-22 \approx 6$  months) and PC3

variogram described an additional shorter-range (j =  $8-9 \approx 2$  months) periodic variation trend. According to the PCA results, all properties were affected by a long-range cyclic variation (loading PC1 and PC2) while iron and pH displayed an additional short-range variation component as they were loading both PC1 and PC3. The shortest cyclic variation of the PC3 variogram confirmed that pH and iron required a higher sampling frequency than other elements/compounds.

The global multivariograms and the multivariogram of the three first PCs scores are shown in Fig. 10, along with the corresponding SDSE profiles.

As before with the HF dataset, the multivariograms of the PCs scores showed a much lower sill compared to the global multivariogram. It resulted in a lower number of increments required to minimize the SDSE of the sampling. Both multivariograms displayed a classical increasing variogram shape. However, the PC1-3 multivariogram captured the long-range periodic variation (j =  $45-47 \approx 11-12$  months) while the raw



Fig. 13. Decision making procedure proposal for water sampling in mining.

data multivariogram captured the noisier, shorter periodic variation of the raw data.

The SDSE profiles associated with both variograms indicate that aggregate sampling reduces the SDSE of the final samples significantly compared to spot sampling. The SDSE profile of the global multivariogram suggested that taking only two increments instead of one reduced the SDSE by almost 55% for all three sampling methods. However, the SDSE remained very high. The SDSE profile indicated that even when 20 increments were collected, the SDSE still higher than 40%. Since the range of global multivariogram was around eight weeks or two months, the sampling frequency must be higher than one every eight weeks. However, the high amount of increments that needed to be collected to minimize the SDSE made such a sampling scheme impractical.

As said, the multivariogram of the first three PCs captured the longrange (yearly variation). Due to this cyclic variation, the required sampling frequency should be at least one every nine weeks, similar to the global variogram's sampling frequency. However, the SDSE profile for the PCs multivariogram suggested that only 10 to 15 increments were needed to reduce the SDSE to less than 10%.

Nevertheless, both PCs and global multivariograms did not capture the six months variation observed in the variogram of certain variables (Fig. 7). The PCA analysis on the univariate variogram of the 12 properties monitored in the weekly data confirmed that the half-year cyclic variation existed and mostly concerned phosphorus, calcium, and iron (Fig. 11). Among them, calcium and phosphorus were directly linked to operating conditions. The half-year cyclic variation of operating-related variables could relate to the different modes operating modes in summer and winter. The projection of the loadings on the PC1 and PC2 plan indicated a clustering trend characterized by three groups of variables: (1) phosphorus, iron, calcium, (2) SPC, sulfate, nickel, and (3) the remaining variable (Fig. 11). The fact that the variograms of phosphorus, calcium, and Iron were highly loaded in the PC2 confirmed that they were similar to each other and different from other variograms. They all possessed a pronounced half-year cyclic variation.

The multivariate variogram for these different groups of variables were computed separately (Fig. 12). The multivariogram of phosphorus, calcium, and iron indicated a range of j = 5 weeks. The short-range of periodic variation had a frequency j = 3-4 weeks while the frequency of the long-range periodic variation is j = 20-25 = 5-6 months. The existence of a short-range frequency suggested that the sampling frequency for those three parameters must be at least one per 4–6 days. The SDSE

plot indicated that an aggregate sample made of ten increments was needed to decrease the SDSE to below 10%. The multivariogram of SPC, sulfate, and nickel was a typical increasing variogram with a range  $j = 15 \approx 3$  months. The SDSE plot suggested that the SDSE would be lower than 10% if eight increments were collected systematically to make the aggregate sample. The multivariogram of the remaining variables was a typical cyclic variogram with a periodic variation of  $j = 45-47 \approx 11-12$  months which indicated a sampling frequency of one every 2.4 months.

#### 4. Discussion

The methodology used in this study to design a sampling protocol for the Kevitsa mine could be generalized as a decision-making procedure which could be further used for other mine sites/mineral processing plants. This procedure is based on a thorough variographic and multivariate data analysis that helps identify the main features in the dataset and other criteria. These can be compared to pre-defined criteria such as sampling objectives, practical, economical, or environmental constraints to define tailored water sampling procedures satisfying all criteria. Precisely, the step-by-step procedure for the general decisionmaking process for water sampling could be (Fig. 13):

- Step 1: Data collection (sampling, probe measurement, historical dataset)
- Step 2: Modeling
  - Preprocessing data for further analysis
  - Basic statistical data analysis of the preprocessed data (range of variation, mean, standard deviation for obtaining variation range of the dataset
  - Variographic approaches:
    - PCA of pre-processed data to investigate the potential correlation between variables, reduce the number of variables and filter noise from the data,
    - Variograms of pre-processed data to obtain the unique individual characteristics of all parameters and designing the optimal sampling protocol for univariable,
    - Variogram of PCA scores to investigate the hidden structures and patterns through variables grouping,
    - Multivariogram of pre-processed data to design a sampling protocol that takes into consideration of the overall variability of the data in one variogram,

- Multivariogram of PCA scores to design a sampling protocol that takes into consideration the relevant component of the overall variability of the data in one variogram,
- Multivariogram of grouped variables sharing similar spatial characteristics to design sampling protocol for different groups of variables.
- Step 3: Criteria definition
  - Define different sampling protocols and the associated characteristics (frequency, SDSE, number of increments, etc.),
  - Defining the pros and cons of each protocol suggested by the previous approaches,
- Step 4: Decision on the final water sampling procedures based on the pre-defined criteria such as sampling objectives, practical constraints, etc., and the criteria defined in step 3. For example, when dealing with a multivariate problem, univariate variogram must not be followed.

In practice, the sampling procedure comprises bottles used in taking increments from the entire stream at different time intervals during the sampling period. However, due to the high instability of water (Miranda-Trevino et al., 2013; Le et al., 2020b), the increments must be stabilized right after sampling, e.g. filtration and freezing. At the end of the sampling period, the increments are aggregated as a final sample sent for analysis. This paper described a procedure to design a water sampling scheme. The next step is to investigate the methodology to preserve the increment so that its properties remain constant until the analysis is performed (Le et al., 2020b) Such a procedure of sampling and preserving water sample should be used in the future as a standard procedure for water quality evaluation on-site both in term of performance improvement and environmental regulations.

#### 5. Conclusions

The variogram analysis was used to study the variability in the water property matrix and to investigate the origin of the variations. The results showed that the variations of water-related parameters can be both correlated and independent, depending on the factors that govern the variations. The range and frequency of variations are not similar for all variables. Unstable variables or operation-related variables can change very fast. Meanwhile, stable and season-related variables can change much slower.

The variogram work confirmed that the sampling frequency for process water must be tailored and customized based on the water composition. The procedure is very mine-specific. While this paper proposed a decision-making protocol applicable to any mine site to ease the sampling design process, the final decision is case-dependent.

The results obtained for the case study investigated confirmed that the SDSE associated with spot sampling is relatively high. Hence, the current practice of spot sampling in the mine is not optimal. However, here, a simple two-increment sample could decrease the SDSE by 60%, and with a four-increment sample by 90%. These values are based on the variability of only five physicochemical properties that could be monitored with high frequency (HF dataset). In reality, other chemical properties should be added to the monitoring program. Therefore, the final SDSE would be higher, and the number of increments needed to decrease the SDSE would increase accordingly.

The weekly sampling dataset could help to design a monthly sampling procedure for environmental control purposes. The results showed that the best sampling choice is to group variables with a similar tendency of variation overtime via PCA analysis on the univariate variograms which allows to group the variables into three major clusters: (1) phosphorus, iron, calcium, (2) SPC, sulfate, nickel, and (3) the remaining variables. For environmental purposes, the multivariogram of phosphorus, iron, and calcium indicates that those variables require a higher sampling frequency compared to other variables (i.e., 4–6 days instead of 2.4 months)."

#### CRediT authorship contribution statement

Thi Minh Khanh Le: Conceptualization, Methodology, Formal analysis, Writing – original draft, Visualization, Investigation, Software. Quentin Dehaine: Software, Validation, Formal analysis, Writing – review & editing. Benjamin Musuku: Resources, Writing – review & editing. Nóra Schreithofer: Supervision, Project administration. Olli Dahl: Supervision.

#### **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.

#### Acknowledgments

The authors would like to thank Boliden Kevitsa mine for industrial support and access to the data. Prof. Kari Heiskanen is acknowledged for his advices, comments, and suggestions. The authors thank Prof. K.H. Esbensen for the very valuable comments provided as part of the pre-examination of TMK Le's PhD manuscript.

#### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

#### Appendix A

See Figs. A1-A3 and Tables A1 and A2.



Fig. A1. Univariate variograms of the ORP of the HF dataset (left). The corresponding SDSE profile (right).



Fig. A2. PCA analysis of the HF dataset. Eigenvalues of the principal components (left), loadings (middle), and scores (right) plot projected on the PC1 vs PC2 component plane.



Fig. A3. (left) original turbidity autosacled data, (middle) de-trended turbidity autoscaled data, (right) experimental variogram of de-trended turbidity data.

Table A1

Summary statistics of the 5 properties from the HF dataset. The coefficient of variation (CV) expressed in percentage.

	SPC (µS/cm)	pH	ORP (mV)	Turbidity (FNU)	DO (% Sat)
Mean Std	3134 16.0	7.8 0.3	48.5 21.2	4.8 0.7	89.8 2.2
CV (%)	0.5	4.4	43.8	14.7	2.5

#### Table A2

Summary statistics of the 12 properties from the weekly dataset.

	SO42-	$S_2O_3^2$	Ν	Р	Ca	Mg	Mn	Ni	Si	Fe	pH	SPC
Mean (mg/l)	607	46	4	0.13	163	71	0.04	0.06	10	0.17	8	2560
Std	77	19	1	0.02	17	10	0.03	0.04	2	0.16	1	196
CV (%)	13	42	31	14	11	14	60	62	20	99	9	8

#### References

- Bezuidenhout, Y., 2009. Implementing an Efficient Mine Water Monitoring and Management Program: Key Considerations. Int. Mine Water Conf. 958–965.
- Boardman, D.I., Glendinning, S., Rogers, C.D.F., 2004. The influences of iron (III) and lead (II) contaminants on lime-stabilised clay. Géotechnique 54 (7), 467–486. https://doi.org/10.1680/geot.2004.54.7.467.
- Bourgault, G., Marcotte, D., 1991. Multivariable variogram and its application to the linear model of coregionalization. Math. Geol. 23 (7), 899–928. https://doi.org/ 10.1007/BF02066732.
- Bro, R., Smilde, A.K., 2014. Principal component analysis. Anal. Methods 6 (9), 2812–2831. https://doi.org/10.1039/c3ay41907j.
- Canada Environment, 2009. Environmental Code of Practice for Metal Mines. Mining Division, Processing Directorate, Resources Sectors Branch, Environmental Stewardship Canada, Canada.
- De Maesschalck, R., Jouan-Rimbaud, D., Massart, D.L., 2000. The Mahalanobis distance. Chemom. Intell. Lab. Syst. 50 (1), 1–18. https://doi.org/10.1016/S0169-7439(99) 00047-7.
- Dehaine, Q., Filippov, L., 2015. A multivariate approach for process variograms. TOS Forum 2013, 169. https://doi.org/10.1255/tosf.76.

- Dehaine, Q., Filippov, L.O., Royer, J.J., 2016. Comparing univariate and multivariate approaches for process variograms: a case study. Chemom. Intell. Lab. Syst. 152, 107–117. https://doi.org/10.1016/j.chemolab.2016.01.016.
- Esbensen, K.H., 2017. Theory of Sampling what's next ? In: Proceedings of the 8th World Conference on Sampling and Blending, pp. 29–38.
- Esbensen, K.H., Friis-Petersen, H.H., Petersen, L., Holm-Nielsen, J.B., Mortensen, P.P., 2007. Representative process sampling — in practice: Variographic analysis and estimation of total sampling errors (TSE). Chemom. Intell. Lab. Syst., Special Issue: WSC-5 2006 Special Issue: WSC-5 2006 88, pp. 41–59. https://doi.org/10.1016/j. chemolab.2006.09.011.
- Fowlie, P., Hart, D., Turle, R., 2001. Guidance Document for the Sampling and Analysis of Metal Mining Effluents: Final report, Ottawa (ON): Environment Canada, Environmental Protection service, Minerals and Metals Division.
- Forssberg, K.S.E., Hallin, M.I., 1988. Process Water Reticulation in a Lead-Zinc Plant and other Sulphide Flotation Plants. In: Sastry, K.V.S., Fuerstenau, M.C. (Eds.), Proc, Symp. Challenges in Mineral Processing. Society of Mining Engineers.
- François-Bongarçon, D., 2004. Theory of sampling and geostatistics: An intimate link. Chemom. Intell. Lab. Syst. 74 (1), 143–148. https://doi.org/10.1016/j. chemolab.2004.03.015.
- C.J. Greet, 2010. CHAPTER 1 The Eureka Mine An Example of How to Identify and Solve Problems in a Flotation Plant, in: Flotation Plant Optimisation: A Metallurgical Guide to Identifying and Solving Problems in Flotation Plants, pp. 1–33.

- Gy, P., 2004. Sampling of discrete materials III. Quantitative approach—sampling of onedimensional objects. Chemom. Intell. Lab. Syst. 74 (1), 25–38. https://doi.org/ 10.1016/j.chemolab.2004.05.015.
- Gy, P., 1998. Sampling for Analytical Purposes, first ed. John Wiley & Sons, Chichester, UK.
- Gy, P., 1992. Sampling of Heterogeneous and Dynamic Material Systems: Theories of Heterogeneity, Sampling and Homogenizing. Elsevier, Amsterdam.
- Gy, P., 1971. L'Echantillonnage des Minerais en Vrac. Revue de l'Industrie Minérale, St Etienne.
- Holmes, R.J., 2004. Correct sampling and measurement—the foundation of accurate metallurgical accounting. Chemom. Intell. Lab. Syst. 74 (1), 71–83. https://doi.org/ 10.1016/j.chemolab.2004.03.019.
- Johnson, N., 2003. Issues in Maximisation of Recycling of Water in a Mineral Processing Plant, in: Water in Mining. Melbourne, pp. 239–246.
- Kinnunen, P., Obenaus-emler, R., Raatikainen, J., Guignot, S., Guimera, J., Ciroth, A., Heiskanen, K., 2021. Review of closed water loops with ore sorting and tailings valorisation for a more sustainable mining industry 278. https://doi.org/10.1016/j. jclepro.2020.123237.
- Kuyucak, N., 2006. Selecting suitable methods for treating mining effluents. In: Water in Mining 2006, Brisbane, Australia.
- Le, T.M.K., Mäkelä, M., Schreithofer, N., Dahl, O., 2020a. A multivariate approach for evaluation and monitoring of water quality in mining and minerals processing industry. Miner. Eng. 157, 106582 https://doi.org/10.1016/j.mineng.2020.106582.
- Le, T.M.K., Schreithofer, N., Dahl, O., 2020b. Challenges in the assessment of mining process water quality. Minerals.
- Levay, G., Smart, R., Skinner, W., 2001. The impact of water quality on flotation performance. In: SAIMM. J. pp. 69–76.
- Liu, W., Moran, C.J., Vink, S., 2013. A review of the effect of water quality on flotation. Miner. Eng. 53, 91–100. https://doi.org/10.1016/J.MINENG.2013.07.011.
- Madrid, Y., Zayas, Z.P., 2007. Water sampling: Traditional methods and new approaches in water sampling strategy. TrAC Trends Anal. Chem. 26 (4), 293–299. https://doi. org/10.1016/j.trac.2007.01.002.
- Minkkinen, P., 2004. Practical applications of sampling theory. Chemom. Intell. Lab. Syst. 74 (1), 85–94. https://doi.org/10.1016/j.chemolab.2004.03.013.
- Minkkinen, P., Esbensen, K.H., 2014. Multivariate variographic versus bilinear data modeling. J. Chemom. 28, 395–410. https://doi.org/10.1002/cem.2514.
- Miranda-Trevino, J.C., Pappoe, M., Hawboldt, K., Bottaro, C., 2013. The importance of thiosalts speciation: review of analytical methods, kinetics, and treatment. Crit. Rev.

Environ. Sci. Technol. 43 (19), 2013–2070. https://doi.org/10.1080/10643389.2012.672047.

Mudd, G.M., 2010. The Environmental sustainability of mining in Australia: key megatrends and looming constraints. Resour. Policy 35 (2), 98–115. https://doi.org/ 10.1016/j.resourpol.2009.12.001.

Muniruzzaman, M., Kauppila, P.M., Karlsson, T., 2018. Water quality prediction of mining waste facilities based on predictive models. Geological Survey of Finland.

- Napier-munn, T., 2014. Statistical Methods for Mineral Engineers: How to Design Experiments and Analyse Data, JKMRC monograph series in mining and mineral processing. Julius Kruttschnitt Mineral Research Centre, Queensland Australia.
- Omo-Irabor, O.O., Olobaniyi, S.B., Oduyemi, K., Akunna, J., 2008. Surface and groundwater water quality assessment using multivariate analytical methods: a case study of the Western Niger Delta, Nigeria. Phys Chem. Earth, Parts A/B/C 33 (8-13), 666–673. https://doi.org/10.1016/j.pce:2008.06.019.
- Opitz, J., Timms, W., 2016. Mine water discharge quality a review of classification frameworks. In: Paul, M., Drebenstedt, C. (Eds.), Proceedings IMWA 2016 Mining Meets Water - Conflicts and Solutions. Freiberg/Germany, pp. 17–26.
- Petersen, L., Esbensen, K.H., 2005. Representative process sampling for reliable data analysis—a tutorial. J. Chemom. 19 (11-12), 625–647. https://doi.org/10.1002/ (ISSN)1099-128X10.1002/cem.v19:11/1210.1002/cem.968.
- Pitard, F.F., 1993. Pierre Gy's Sampling Theory and Sampling Practice, Second Edition: Heterogeneity, Sampling Correctness, and Statistical Process Control. CRC Press, Boca Raton.
- Saunders, I.W., Robinson, G.K., Lwin, T., Holmes, R.J., 1989. A simplified variogram method for determining the estimation error variance in sampling from a continuous stream. Int. J. Miner. Process. 25 (3-4), 175–198. https://doi.org/10.1016/0301-7516(89)90017-3.
- Schumann, R., Levay, G., Ametov, I., 2009. Impact of recycling on process water quality in mineral processing. In: Proceedings Water in Mining 2009. The Australasian Institute of Mining and Metallurgy, Melbourne, pp. 79–86.
- Whaley-Martin, K., Marshall, S., Nelson, T.E.C., Twible, L., Jarolimek, C.V., King, J.J., Apte, S.C., Warren, L.A., 2020. A mass-balance tool for monitoring potential dissolved sulfur oxidation risks in mining impacted waters. Mine Water Environ. https://doi.org/10.1007/s10230-020-00671-0.
- Wise, B.M., Gallagher, N.B., 1996. The process chemometrics approach to process monitoring and fault detection. J. Process Control 6 (6), 329–348. https://doi.org/ 10.1016/0959-1524(96)00009-1.
- Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemom. Intell. Lab. Syst. 2 (1-3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9.