Mine Water and the Environment (2019) 38:353–365
https://doi.org/10.1007/s10230-019-00598-1

TECHNICAL ARTICLE

Experimental Evaluation of Surface Water Sampling Variability for Environmental Monitoring in Iron Ore Operations

Karin Engström[1,2] · **Liselotte Olausson**[3] · **Kim H. Esbensen**[2,4,5]

Received: 2 May 2018 / Accepted: 26 February 2019 / Published online: 22 March 2019

Abstract

Environmental self-monitoring is a government requirement for Swedish process industries. This includes sampling and analysis of recipient water that might be adversely affected by emissions. The requirements for accredited analytical methods are strict, with well-defined measurement uncertainties, but estimations of the attendant sampling variability are seldom required, presented, or evaluated in environmental surface water sampling. The goal of this study was to perform an initial evaluation of the measurement variability for surface water sampling within the self-monitoring program for a large mining company in northern Sweden. The results indicate that the method for evaluation of sampling and measurement variability itself affects the results obtained. Therefore, the evaluation scope must be clearly defined in advance, so that the most appropriate approach, resulting in a realistic quantification of total variability, can be selected. This study shows that duplicate sampling experiments result in significantly larger sampling variability estimates when accounting for ambiguities in the sampling protocol than similar experiments under repeatability conditions. This is due to large temporal variations in stream flux and analyte concentrations in the evaluated sampling targets. The ambiguities in different sampling protocols must be fully described and considered when designing empirical evaluation experiments to allow valid evaluation of the total sampling and measurement system variability. An automated sampler using volume-proportional sampling to collect increments for composite samples is recommended to reduce unnecessary sampling variability and address significant temporal changes in stream flux and analyte concentrations appropriately.

Keywords Sampling · Recipient water · Duplicate sampling · Replication experiment · Theory of sampling · TOS

Introduction

The Theory of Sampling (TOS) is a complete theory covering all aspects of sampling particulate material (Esbensen 2016; Gy 1979). Even though TOS was originally developed for sampling of particulate materials, all main principles related to sampling errors and heterogeneity can be applied to practically any form of material sampling, including surface water (Gy 2004). Environmental self-monitoring is mandatory in most countries for industries that might affect the environment adversely. The company must control the impact on the environment through self-monitoring, which should be internally developed based on the operations at hand. Even though some countries require mandatory education of water sampling staff, the sampling methods used before analysis are often not in compliance with TOS. In surface water sampling, as in any other sampling situation, the heterogeneity of the target lot will influence the validity of the sampled water/material. It is essential to consider and counteract the lot heterogeneity in any sampling protocol. This is done by collecting equi-probabilistic samples where all parts of the lot must have an equal probability to end up in the sample (Gy 1998, 2004; Pitard 1993). Only by following this fundamental sampling principle (FSP), and all

---

✉ Karin Engström
karin.engstrom@lkab.com

[1] Sampling and Test Methods, Luossavaara-Kiirunavaara AB (LKAB), Kiruna, Sweden

[2] Aalborg University, Aalborg, Denmark

[3] Luossavaara-Kiirunavaara AB (LKAB), Kiruna, Sweden

[4] Geological Survey of Denmark and Greenland (GEUS), Copenhagen, Denmark

[5] Université du Québec à Chicoutimi, Chicoutimi, QC, Denmark

complementary principles laid out by TOS, can a documented representative sample ever be collected (Esbensen and Wagner 2014).

Analytical results will always be accompanied by uncertainty arising from remaining, or insufficiently reduced sampling errors, in addition to analytical uncertainties. Sampling errors are divided into two categories: correct sampling errors (CSE) and incorrect sampling errors (ISE). The CSE include fundamental sampling errors (FSE) and grouping and segregation errors (GSE), which are always present in sampling situations, but can be minimized using correct sampling protocols. The ISE, incorrect delimitation errors (IDE) and incorrect extraction errors (IEE) relate to the geometrical delineation and physical collection of the sample (increment), while incorrect preparation errors (IPE) relate to the post-sampling processes itself, including sample preparation and possible contamination during the activities carried out on the sample between primary sampling, sub-sampling (mass reduction), and chemical analysis (Esbensen and Paasch-Mortensen 2010). The ISE must be eliminated in order to be able to guarantee a bias-free sampling process (Danish Standard, DS 3077 2013).

Even though it is claimed to be widely agreed that the measurement process starts when the primary sample is collected from the lot, the analytical uncertainty of chemical measurements is traditionally the main focus of the measurement uncertainty ($\mathrm{MU}_{\mathrm{analysis}}$) of reported analytical results (Ramsey et al. 2011). However, the complete measurement uncertainty $\mathrm{MU}_{\mathrm{sampling+analysis}}$, including the variability stemming from sampling (primary and sub-sampling), is still seldom presented or even evaluated in environmental surface water sampling (Esbensen and Wagner 2016).

Several recent publications have acknowledged this lack of focus on sampling variability in environmental monitoring, presenting examples of experimental methods used for estimating both sampling and analytical uncertainties in various environmental applications (Botta et al. 2012; Guigues et al. 2016; Ramsey 1998; Ramsey et al. 2011; Rode and Suhr 2007). The increased scientific understanding and focus on the importance of sampling variability in environmental control programs can be expected to influence the governmental control of mining operations and other industries affecting air and water emissions. This prospect is an incentive for the industry to act proactively by including control of sampling variability in the mandatory self-monitoring programmes related to environmental analysis. It has long been mandatory to evaluate and document analytical measurement uncertainty, including any contribution from analytical bias within the laboratory, together with any accredited analytical result. However, accounting for the uncertainty stemming from primary sampling in the field and secondary sampling operations in the laboratory is not generally expected in environmental monitoring. This paper points out the consequences that will result if this status quo continues unchanged.

Environmental Monitoring and Governmental Regulations in Sweden

The Luossavaara-Kiirunavaara limited company (LKAB) is a state-owned mining company in northern Sweden. The core business is producing high quality iron ore pellets for blast furnace and direct reduction steel making. According to governmental regulations in Sweden, the operator must provide the knowledge and evidence of the environmental impact attributed to the operation. This is done through self-monitoring, which is a continuous process usually involving sampling, analysis, studies, and investigations. If the environmental permit is associated with specific terms, these are to be controlled by the self-monitoring and the results are reported to the authorities. How to perform, follow up, report, and document the self-monitoring is described in a control program, which normally includes sampling protocols. These protocols also describe sampling targets and frequencies, as well as the necessary requirements for analytical parameters and analytical accreditation. Monitoring and control of the analytical uncertainty for each accredited analysis is closely controlled by accreditation officials. However, the sampling variability coupled with the specific surface water sampling has not been evaluated at LKAB previously and directed experiments were therefore critically needed, especially considering that total sampling error effects often dominate total uncertainty budgets by factors of 5–25, or more. This paper summarizes the first experiments to advance in this direction.

Current Practice for Water Sampling at LKAB

Water sampling at LKAB is based on the sampling methods advised by the Swedish environmental authorities. The scope of the control program (e.g. frequency of sampling) used for self-monitoring varies between mining sites, as this is governed by the individual permits issued for each site. The environmental permit for LKAB operations in Svappavaara stipulates that two surface water samples shall be collected and analysed per week from the sampling target SVA (Svappavaara water). Even though specific days for sampling are not regulated in the permit, the sampling is normally performed every Tuesday and Thursday for practical reasons, such as sample transportation and laboratory hours. For the LKAB operations in Kiruna, the environmental permit stipulates extraction and analysis of one surface water sample per month from the sampling target KVA (Kiruna water). The sampling protocol specifies the week that the sampling will take place, but not a specific day. Most of the self-monitoring surface water samples at LKAB are planned annually on a weekly or monthly basis, as with KVA. This gives the sampling technicians some freedom to plan their working hours according to other tasks as well as favourable weather conditions.

Both sampling targets, KVA and SVA, are located in smaller streams, approximately 10–15 m wide. The KVA sampling target is at the beginning of an artificial dike that serves as the outlet from the clarification pond, leading the water to the lake Mettä Rakkurijärvi, the recipient water system (Fig. 1a). Approximately 98% of the water in KVA is discharged from the clarification pond, while only 2% is natural flow from precipitation and ground water. The SVA sampling target is in the natural stream, Liukattijoki, which periodically receives discharge water from the water equalising storage (Metträsket), which in turn receives water from the tailings dam via the clarification pond (the water is pumped in pipelines from the clarification pond to Metträsket; Fig. 1b). The discharge from Metträsket to Liukattijoki varies between 0 and 50% relative to the natural flow in the stream. The discharge can be momentarily high, but no continuous outflow is present, meaning that the concentrations at SVA is mainly related to the level of discharge relative to the natural flow in Liukattijoki. The highest ratio of discharge water occurs when the natural flow is low and a large release from Metträsket is initiated. This means that the absolute analyte concentrations in the clarification pond does not affect the concentrations at SVA to the same degree as it does at KVA. In Kiruna, the discharge from the

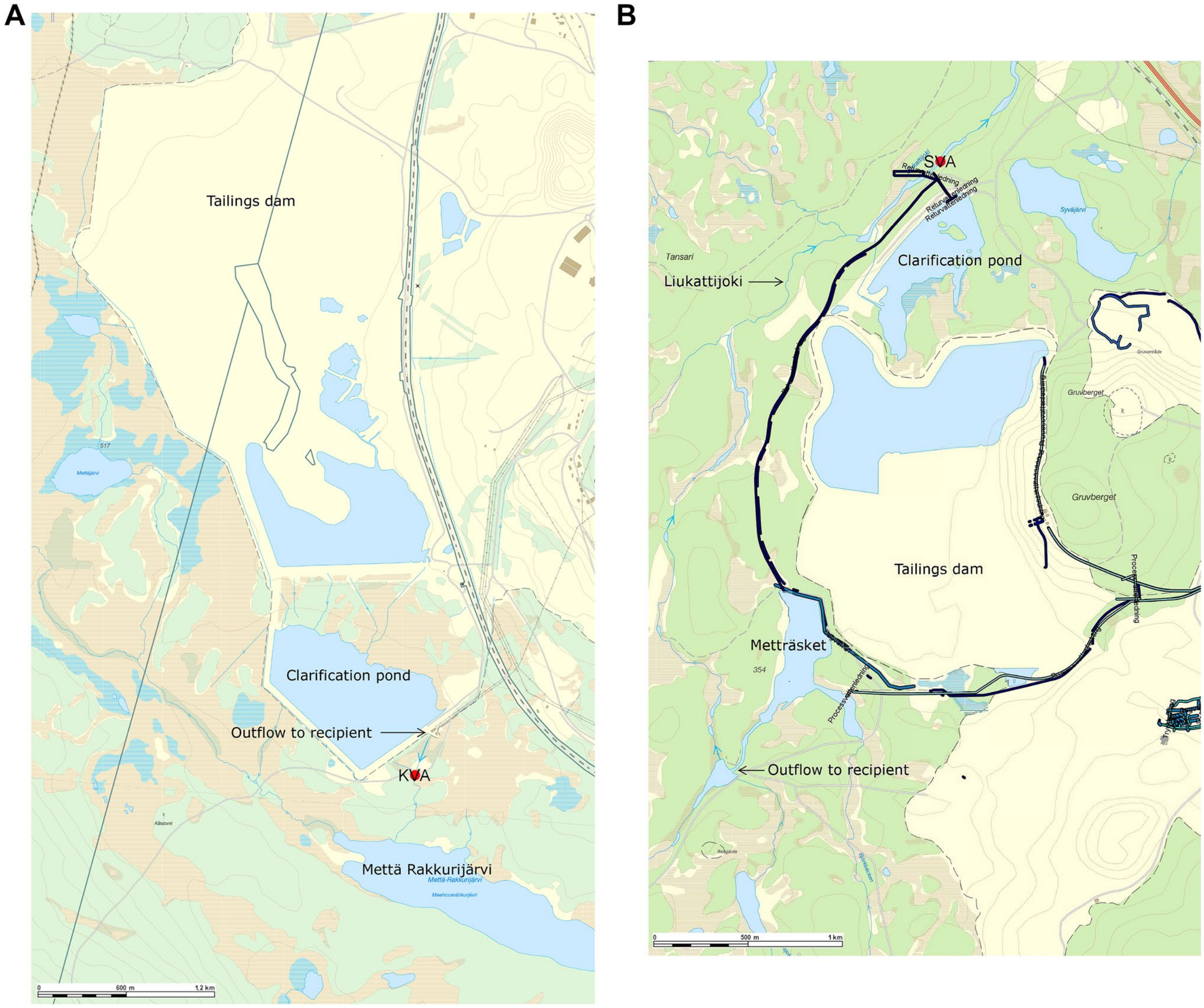

Fig. 1 Water systems in **a** Kiruna and **b** Svappavaara, with tailings dams, clarifications ponds, Metträsket, Mettä Rakkurijärvi, outflow location and sampling targets KVA and SVA annotated. The black lines in **a** corresponds to the pipelines pumping water from e.g. the clarification pond to Metträsket

clarification pond to the artificial dike is more uniform over time, as it is the major part, ≈ 98%, of the stream water at the KVA sampling target.

The objective of the measurement systems at both SVA and KVA is to monitor the analyte concentrations released to the recipient streams. The goal is therefore not to identify peak concentrations, but rather to monitor the mean concentration (and total amount) being released from the process to the recipient water system. The environmental permits stipulate the maximum allowable concentrations reached in the recipient streams, rather than the concentrations being released, meaning that monitoring of natural flow in the recipient is an important complement to the chemical analysis of the sampled water.

General and Local Variability in Recipient Stream Water

Diurnal, seasonal, and event-driven variability of water quality is present in any water catchment. These variations are, for example, dependent on air-temperature, solar radiation, photosynthesis, respiration, stream flow variation, circulation, and photochemical processes (Henjum et al. 2010). Bio-geochemical processes can be affected by the solar photo-cycle and the level of diurnal scale variation can be as large as annual or seasonal time scales. The 24-h variations in metal concentrations have also been observed to be larger than long term, e.g. seasonal variations (Nimick et al. 2005). Metal concentrations in stream water can vary in relation to the downstream distance from the primary source and has been shown to behave differently, depending on the contaminant being studied. As the diurnal and groundwater variabilities also is significant for metal contaminants, the relationship of these variabilities needs to be addressed (Gammons et al. 2007).

The size of the clarification pond at Kiruna is 2.3 mm^3, and the process water inflow to the tailings dams and clarification ponds is close to uniform over time at both Kiruna and Svappavaara. The changes in inflow of water is therefore mainly affected by snow melting and precipitation. The seasonal variability of natural flow in the recipient water streams is large and similarly dependent on snow melting and precipitation. Flow peaks are generally registered during the spring flood and during heavy rains in the summer, while the snow-covered period in winter leads to significantly lower flows (Westerberg and Huseby Karlsen 2017). At Kiruna, increased inflow to the clarification pond leads to a direct increase of outflow to the recipient water system, as the overflow is controlled by a fixed threshold. The clarification pond at Svappavaara (0.3 m^3) is significantly smaller than at Kiruna and is therefore followed by an equalising water storage pond, Metträsket, for additional storage capacity. The outflow from the clarification pond runs via Metträsket and is controlled relative to the water level there. This system leads to an uneven outflow from the internal water system to the recipient stream, Liukattijoki, as hatches are opened to release larger amounts of water when the water level in Metträsket reaches a predetermined set point. This results in momentarily large water emissions, rather than a regular outflow over longer time periods. Therefore, concentration variations in Liukattijoki largely depends on the outflow flux from the water storage relative to the natural flow in the stream, which in turn exhibits large seasonal variations, from ≈ 0.2 to ≈ 8 m^3/s (Westerberg and Huseby Karlsen 2017). An important limitation for the control of the outflow at both Kiruna and Svappavaara is the restricted storage capacity in the clarification ponds and Metträsket. A regulated overflow, related to the natural flow in the recipient streams, would be preferred to even out the analytes concentrations in the recipient water system. However, this is currently not possible as the overflow must be regulated according to the storage capacity in the ponds.

The circulation in the LKAB clarification ponds is continuously high due to the large inflow of process water via the tailings dam. This is coupled with water being pumped back to the processing plants from the clarification pond outflow. The high circulation in the water system means that there is limited time for sedimentation in the clarification pond. As the circulation is mainly driven by the exchange of process water, the effects of ice coverage and wind is negligible in comparison.

The analyte concentrations in the discharged water from the clarification ponds at Kiruna and Svappavaara are not directly related to the viability of water inflow and level of dilution. For some analytes, e.g. nitrogen, dilution is visible during high precipitation. However, for sulphate, no direct dilution is visible as chemical equilibrium exists in the clarification pond. This leads to sulphate being dissolved from the sediments into the water if increased precipitation lowers sulphate concentrations in the pond. Thus, the sulphate concentration is not directly affected by dilution during high precipitation events. However, the dilution factor for sulphate depends on the water temperature, resulting in lower sulphate concentrations in the winter (with water temperatures close to 0 °C) than in summer (Ylipää and Björnfot 2018). Another reason for seasonal variation is biochemical process in the clarification pond, e.g. bacterial denitrification, which increases during the warm summer season with increased sunlight. Internal calculations indicate that there is approximately a 20% decrease in nitrate emissions to the Kiruna recipient water due to denitrification during the summer months (Ylipää and Björnfot 2018).

Methods for Evaluating Sampling Variability

Several methods, using both empirical and modelling approaches for evaluation of sampling and analytical uncertainties in environmental monitoring have been published recently (Guigues et al. 2016; Lyn et al. 2007; Ramsey et al. 2011). One recommended approach is the duplicate method, which is both simple and cost-effective. The approach of the duplicate method is that the field sampler collects duplicate primary samples from 10% (but no less than eight samples) of the sampling targets (Grøn et al. 2007). To enable evaluation of the complete uncertainty, including primary sampling, possible sub-sampling, and sample preparation, it is vital to duplicate the primary sample extraction, rather than just duplicating the sample preparation or sample analysis.

According to Ramsey et al. (2011) one important limitation of the duplicate method is that it excludes the uncertainty arising from systematic sampling bias effects. A method for inclusion of any so-called systemic effects in sampling is the sampling proficiency test (SPT). This method involves at least eight experienced samplers sampling from more than five typical sampling targets (Ramsey et al. 2011). However, an SPT using several samplers is expensive and often impossible to carry out due to proprietary reasons. It is therefore seldom practicable for validation of a sampling protocol in a single organization (Ramsey and Thompson 2007). The SPT approach has also been criticised by Esbensen and Wagner (2014), based on conceptual and theoretical problems with this approach; the main critique is the unrealistic concept of a "constant sampling bias" (Esbensen and Wagner 2014). Within TOS, it has been shown that there does not exist a constant sampling bias or a systematic sampling effect (breaking with the oft-assumed analogy to the analytical bias). In fact, every time a sampling bias is estimated, it will be different, i.e. the sampling bias is inconstant and variable, because of irregular lot heterogeneity. The sampling bias (always random) must be reduced, preferentially eliminated, by specific application of all necessary elements in TOS that are aimed at achieving correct sampling. On this basis, i.e. when a non-biased sampling process has been achieved, the use of duplicate and/or replication experiments will then be sufficient and effective for estimating the complete sampling variability, as manifested by $\mathrm{MU}_{\mathrm{sampling+analysis}}$ (Esbensen et al. 2007; Esbensen and Wagner 2016; Pitard 1993). As the aim of this paper is to apply the concepts of TOS to environmental recipient water sampling, SPT is not applicable, but correct duplicate and replication experiments are sufficient for estimating all sampling error effects.

An important aspect when assessing the variability in sampling-and-analysis is setting an acceptance level below which one can claim a "fit-for-purpose" status, i.e. what is the relative measurement uncertainty, including an acceptable variability stemming from both sampling and analysis for receiving meaningful data that can be used for evaluation and decision making. Even though such thresholds need to be set relative to the specific lot material and sampling procedure used, Ramsey et al. (1992) suggests that the total measurement variability (including sampling) should not contribute more than 20% of the total variability of the analyte over time. In the sampling community, a general consensus indicates that a threshold of 20% for the relative sampling variability (i.e. the sampling variability relative to the average analyte concentration) is also recommended here (Esbensen and Wagner 2016). However, TOS also strongly advocates individual assessment and determination of sampling variability thresholds, for each specific sampling procedure, since sampling targets generally display significantly different heterogeneities (DS3077 2013).

Study Objectives

The objectives of this study were to perform an initial evaluation of the sampling variability associated with the current LKAB environmental measurement system for emissions to surface water. The sampling systems at the sampling targets KVA and SVA were evaluated using the duplicate method and replication experiments. For all reported experiments, the replication/duplication was executed at the point of primary sample extraction so as to include all variability effects stemming from the complete sampling to analysis pathway. A second objective was to evaluate the experimental methods for estimating sampling variability in recipient water sampling for environmental monitoring.

As the objective was to conduct initial evaluations of the sampling variability, together with evaluation of the specific experimental methods, the time frames for the experimental campaigns were limited. One consequence of this is that the complete seasonal variability in analyte concentrations and flow were not covered. However, the study will provide recommendations for how to implement comprehensive and continuous control of the sampling variability of surface water sampling at LKAB. Note that the study did not evaluate the seasonal or yearly variability of analyte concentrations in the stream water but was limited to studying the specific sampling variabilities of the current measurement systems.

Sampling and Analytical Methods

All sampling experiments were conducted on the sampling targets regularly used in the current self-monitoring programs at LKAB. The KVA and SVA sampling targets were selected as they are an important component of the environmental permits and have frequent historical data for all important analytical parameters. For the KVA sampling target, one duplicate- and two replication sampling experiment were conducted. For the SVA sampling target, only one duplicate- and one replication sampling experiment were conducted.

The current physical extraction of water samples method used by LKAB was used during all experiments in this study. Within the TOS understanding, this approach is a manual grab sampling (usually rejected, as it can never be representative), which is the official method described and taught by water sampling consultants that follow the Swedish government regulations for environmental self-monitoring. The sample extraction is to be performed in the centre of the targeted stream, approximately 50 cm from the surface (or at mid depth if the stream is shallower than 1 m). To be able to reach the centre of the stream, a sample container can be attached to an extension pole (Fig. 2). To allow for a sufficient amount of water to be collected, three increments are collected from the same location and combined into one primary sample. The current grab sampling introduces IDE and IEE sampling errors as the sample extraction does not allow equal probability for the complete lot (flux cross section) to end up in the sample. A TOS correct sample extraction requires that the complete cross section of the lot be covered by the sampling equipment and that the speed with which it is crossing from one side of the stream to the other is uniform. Hence, if the volume or speed of water is less in the shallower water at the sides of the stream, the amount of water collected should be less than in the centre of the stream where the flow is faster. Currently, it is impossible to achieve this proportional, TOS-correct sample extraction using the recommended sampling container.

In an attempt to handle some of the ISE associated with the traditional grab sampling of stream water, Ramsey (2014) presented the Aloha Sampler™. The design of this sampling equipment (Fig. 3), enables sample collection from the complete cross section of a reachable water stream, as is the case, for example, at sampling target KVA. The small holes in the lid of the sampler allow for water to enter and an equivalent volume of air to exit in a controlled, slow fashion (the Aloha Sampler™ comes with optional hole diameters), enabling sampling staff to move the sampler and collect water across the stream in a continuous motion (Ramsey 2014). The Aloha Sampler™ was used in one replication experiment at the KVA sampling target for comparison with the current sample extraction method used by LKAB. The continuous sampling motion applied in this study, using the Aloha Sampler™, is visualised in Fig. 4.

All water samples were analysed at the LKAB environmental laboratory in Kiruna or at ALS in Luleå, using analytical methods accredited by official national accreditation agencies to ensure valid analyses for the environmental permits. A list of analytical methods and locations for analysis is presented in Table 1. Note that samples are routinely analysed for several additional parameters. The ones presented in this study were selected because they are coupled with possible ecological effects in the recipient water and are therefore important for the LKAB environmental permits.

Fig. 2 Extendable pole with standard plastic sample container used to collect water samples according to the official protocols in the present study

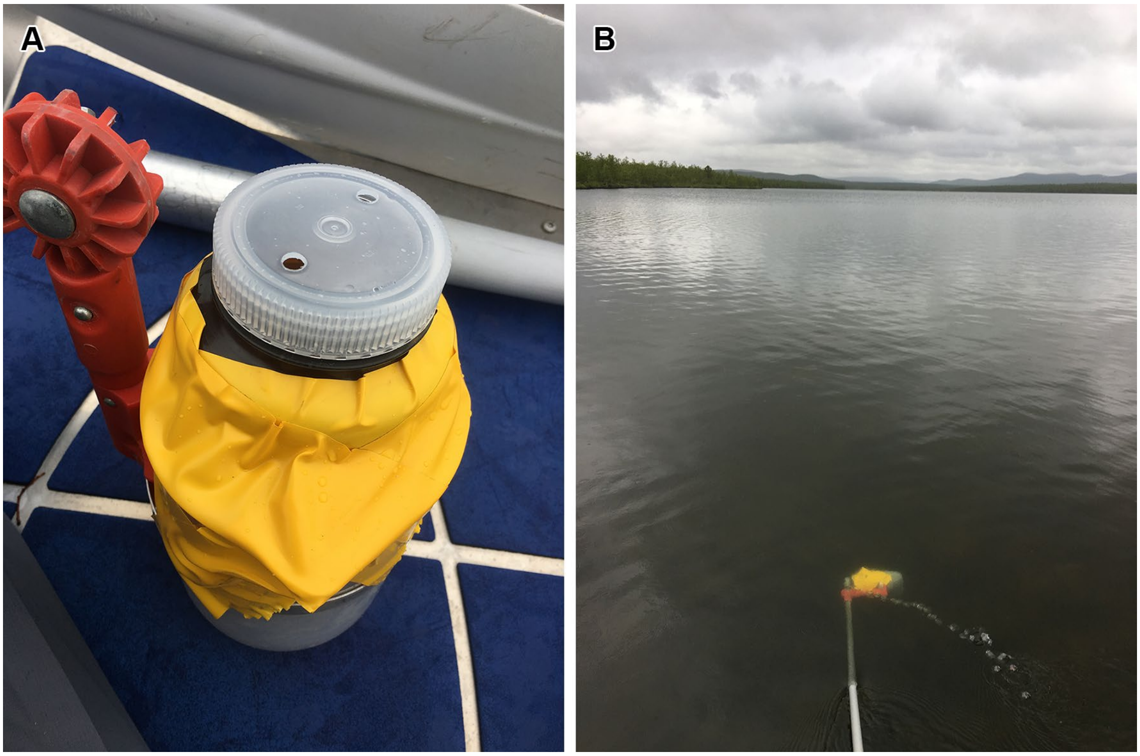

Fig. 3 The Aloha Sampler™ designed for composite sampling, for example along a complete cross section of the streaming water

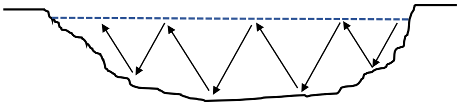

Fig. 4 Aloha Sampler™ used in a continuous motion to collect water from a complete cross section of a stream (Ramsey 2014)

Duplicate Sampling Experiment

The duplicate method is the most straightforward method for assessing variability stemming from sampling, sub-sampling, and analysis relative to overall data variability (Ramsey 1998). According to Lyn et al. (2007), duplicates should

Table 1 Location for analysis of the evaluated analytical parameters

	Analytical method	LKAB laboratory Kiruna	External laboratory
Alkalinity	ISO 9963-2 Water quality—determination of alkalinity—part 2: determination of carbonate alkalinity	X	
Conductivity	SS-EN 27 888 Water quality—determination of electrical conductivity (ISO 7888:1985)	X	
Total phosphorus	SS-EN ISO 6878:2005 Water quality—determination of phosphorus—ammonium molybdate spectrometric method (ISO 6878:2004)	X	
Nitrate sulphate	SS-EN ISO 10304-1:2009 Water quality—Determination of dissolved anions by liquid chromatography of ions—part 1: determination of bromide, chloride, fluoride, nitrate, nitrite, phosphate and sulphate (ISO 10304-1:2007)		X
Copper Vanadium Zinc Uranium	SS EN ISO 17294-1, 2 (mod)/EPA-method 200.8 (mod) Water quality—application of inductively coupled plasma mass spectrometry (ICP-MS)		X

be collected in accordance to the current sampling protocol, reflecting all ambiguity that can be expected. This means that duplicates are separated in time and/or space by the same ambiguity that is presented by the sampling protocol. The ambiguity of a sampling protocol leads to diverse interpretation or application of the protocol in different situations, e.g. variations in the sample extraction location or variations in the time for sample extraction. The primary sample duplicates should ideally also be analysed in duplicate to enable the use of analysis of variance (ANOVA) for separation and quantification of the variability stemming from sampling and analysis, respectively (Lyn et al. 2007). It is important to clearly describe the exact step where duplication is applied to allow clear indication of which variabilities are included in each variance estimation. To accommodate for possible outlying values, the robust ANOVA (RANOVA) is the recommended statistical method, as the classic ANOVA is sensitive to outlying values due to sampling, geochemical, or technical issues (Ramsey et al. 1992). The robust method is claimed to provide valid estimates of the variability components 'under average background levels'. For example, single outliers do not affect the results of a RANOVA unduly, like it would classic ANOVA (Ramsey et al. 1992). There are serious misgivings in the TOS community regarding this position regarding outliers.

Two duplicate sampling experiments were carried out at the KVA and SVA sampling targets. The KVA duplicate sampling experiment was carried out in accordance to the method described by Lyn et al. (2007), which stipulates that the distance between the duplicated samples in time or space should be separated according to the ambiguity in the sampling protocol. Especially for sampling targets with large temporal variability, e.g. stream water, as is the case for both KVA and SVA, the periodic separation between duplicates should correspond to the uncertainty in time that is described by the sampling protocol (Ramsey 1998). In this case, the KVA sampling protocol specify a fixed week for sampling once every month, while the day for sampling is open within that week. This means that the ambiguity in the KVA sampling protocol can lead to a temporal spacing between routine samples of up to 6 days (within 1 week). To reflect the possible variability within the sampling protocol, this sampling experiment extracted duplicates with random time spacing between 1 and 6 days. Each duplicate sample was analysed in duplicate for the important parameters, including alkalinity, conductivity, phosphorus, nitrate, sulphate, copper, vanadium, zinc and uranium. The experiment was conducted during 8 consecutive weeks, resulting in a total of 8 duplicate samples and 32 individual assay results per parameter.

As SVA is sampled twice a week, routinely on specific days, the sampling protocol is more precise than at KVA. The SVA experiment was carried out by extracting duplicate samples under repeatability conditions, according to the definition in Grøn et al. (2007, p. 29): samples are collected at the same time, by the same person, at the same sampling target, using the same sampling equipment. This minimizes variability introduced by primary sample extraction, excluding effects related to long-term temporal changes in the quality of the sampled water, or changes in sampling staff or sample equipment.

For the SVA duplicate samples, no duplicate analysis was conducted. However, as the samples go through identical sample preparation and laboratory analysis at the same laboratories as the KVA samples, the analytical uncertainty were assumed equal to these samples. The duplicate samples were collected once a week (at every other routine sampling extraction) for 15 weeks, resulting in 15 paired samples.

Replication Sampling Experiment

Replication experiments (RE) are used to calculate the relative sampling variability (RSV), a sampling quality index describing the variability for a specific sampling and analysis procedure (DS3077 2013). The RE is particularly useful when sampling stationary lots, in comparison to variographic characterization, which is more powerful when assessing measurement systems performance for dynamic lots in process control (Esbensen and Wagner 2016). The RE is a practical and informative method for both initial and regular quality assessment of recipient water sampling.

In this study, the sampling of SVA targets were replicated using the current sample extraction method and for KVA, both the current sample extraction and the Aloha Sampler™ were used in two separate RE. During the experiment, ten replicate primary samples were collected in direct repetition and brought to the laboratory for analysis. The objective for not accounting for the temporal ambiguity in the sampling protocols was to enable direct comparison of the results from the two separate sampling targets (KVA/SVA) and the two alternative sampling methods (current/Aloha Sampler™). The complete primary sampling was replicated, meaning that for each of the ten replicates, three primary increments were collected and combined into one primary composite sample. For one of the replicate samples, the analysis was also replicated ten times for the parameters alkalinity, conductivity, and total phosphorus. This allows for estimation of both the RSV as well as the isolated analytical variability under repeatability conditions.

Expression of Results

The relative standard deviation for sampling (s_s) and analysis (s_a) respectively, as well as the total variability for the measurement system (s_m) was calculated for the evaluated sampling targets, Eq. 1. It is common to present uncertainties as a relative expanded uncertainty (U), with a coverage factor of k = 2, Eq. 2 (Ramsey 1998). Under the assumption of normally distributed data, this results in an uncertainty that describes a 95% confidence interval for the average result ($\bar{x}$). In this study, results will be presented as relative expanded uncertainties, denoted as U, with a uniform coverage factor of 2. The relative uncertainty is especially useful for comparing different analytes and different experimental methods:

$$s_m = \sqrt{{s_s}^2 + {s_a}^2}, \tag{1}$$

$$U\% = 100\frac{2s}{\bar{x}}\ \%. \tag{2}$$

Results

Duplicate Sampling Experiment

The results from the two separate duplication experiments showed different sampling variability. The variabilities observed for KVA were significantly higher than at SVA for all analysed parameters (Table 2). The absolute concentrations for all analytes were also higher at KVA than for SVA. This is because the water at KVA is mainly water from the clarification pond, while SVA is a pre-existing stream with mostly natural water and only temporary overflow from the clarification pond. The water in the clarification pond preceding KVA also has higher concentrations of most elements as the Kiruna ore is richer in salts and the processing water is more recirculated in the Kiruna processing plants. Note that the expanded relative uncertainty from sampling, for sampling target SVA, is calculated from the total measurement uncertainty for SVA and the uncertainty from analysis for KVA sampling target (as no duplicate analysis was conducted for SVA). This will have no practical implications on the evaluation as the sampling preparation and analytical process is identical for both sampling targets (i.e. the primary samples are delivered to the same laboratory and are prepared and analysed by the same personnel, using identical methods and equipment for all incoming water samples).

The results from the duplicate experiment show that the measurement variability (including sampling), relative to the total variability, was less than 20% for SVA, which is within recommended limits for this sampling target. For KVA, the total measurement variability (including sampling), relative to the total variability, ranged from 15 to 80% for the analysed parameters. Alkalinity, conductivity and phosphorus were within the requirement of less than 20%, while the variability for the remaining parameters exceeded this threshold. The main reason for the differences is that the duplicates at SVA were collected under repeatability conditions (as a reflection of the smaller ambiguity in the sampling protocol for SVA), while the temporal separation of duplicates at KVA was up to 6 days (as a reflection of the larger ambiguity in the sampling protocol for KVA). The temporal ambiguity in the sampling protocol for KVA, translated to the duplicate experiment, means that the variations in absolute analyte concentrations in the stream water, over 6 days, were included in the measurement system variability assessment. This is a direct consequence of the KVA sampling protocol that allows a large ambiguity in the time for sample extraction. Comparatively, the sampling protocol for SVA has a more strict time spacing, where two samples are collected every week and routinely collected the same weekdays, i.e. the temporal ambiguity is significantly less for the SVA sampling protocol and the measurement variability relative to the total variability is therefore also significantly lower.

Replication Sampling Experiment

The results from the three replication experiments (SVA, KVA, KVA_{AS}) indicate that the relative measurement variabilities (including sampling and analysis) for all of the analysed parameters were below the 20% threshold (Table 3). For the three common parameters (conductivity, nitrate, and sulphate), both sampling targets show similar results for the traditional sampling method used by LKAB, despite the large difference in absolute concentrations. The remaining six parameters were only analysed for one of the two sampling targets as a direct reflection of the two separate regulatory sampling protocols requiring different parameters.

However, the duplicate experiment at KVA showed that the large temporal variability in the stream water increased measurement variability significantly when the duplicates were spaced further apart. Furthermore, the results from the Aloha Sampler™ replication experiment (KVA_{AS}), Table 3, showed slightly higher total measurement variability than the traditional sample extraction method. This could be due, in part, to the continuous motion of the Aloha Sampler™ collecting an entire cross section of the stream, which allows for a wider coverage of the overall heterogeneity in the streaming water. Another possible influence for the Aloha Sampler™ procedure is its up-and-down motion, which results in a differential influence due to the hydrostatic pressure in the stream; more water will be forced into the sampling container as depth increases. If the evaluated analyte is unevenly distributed with depth, this could therefore result in higher variability for the Aloha Sampler™ approach relative to stationary grab sampling. The increment collection is also slower than for the traditional grab sampling from the centre of the stream. This means that the time frame for collecting all ten replicate samples increased from ≈ 15 min with the traditional method to 1 h for the Aloha Sampler™. This

Table 2 Variability results for duplicate sampling experiments for SVA and KVA sampling targets

Chemical parameter	Unit	$U_{a\text{-lab}}$	$U_{a\text{-dupl}}$	$U_{s\text{-dup}}$		$U_{m\text{-dup}}$		% meas var		Mean	
		Expanded relative uncertainty from analysis (incl. bias) (95%) (from laboratory QA/QC program) (%)	Expanded relative uncertainty from analysis (95%) RANOVA of duplicates (%)	Expanded relative uncertainty from Sampling (95%) RANOVA of duplicates (%)		Expanded relative uncertainty from measurement (95%) (sampling + analysis) RANOVA of duplicates (%)		% of total variance for measurement (sampling + analysis) RANOVA of duplicates (%)		Mean value/ concentration during experiment	
			KVA	SVA	KVA	SVA	KVA	SVA	KVA	SVA	KVA
Alkalinity	mmol/l	10	0.0	–	3.2	–	3.2	–	20.9	–	48.7
Conductivity	mS/m	5	0.5	0.2	0.6	0.6	0.8	0	13.7	23.3	284.1
P_{tot}	mg/l	25	7.6	–	18.1	–	19.6	–	14.9	–	0.028
Nitrate	mg/l	15	3.1	< 0	8.7	1.5	9.2	0.01	36.1	0.70	19.41
Sulphate	mg/l	15	6.8	< 0	28.0	2.0	28.8	0.02	79.6	60.3	1369
Copper	µg/l	≈ 20	8.2	15.6	34.1	17.6	35.1	10.4	67.1	0.60	1.75
Vanadium	µg/l	≈ 20	4.0	12.3	27.2	13.0	27.5	3.0	49.4	0.12	2.49
Zink	µg/l	≈ 20	32.3	32.0	58.4	45.5	66.7	8.6	30.0	2.48	3.32
Uranium	µg/l	≈ 20	6.0	< 0	17.8	5.0	18.8	0.18	31.8	0.72	13.1

Table 3 Variability results for replication sampling experiments for SVA and KVA sampling targets

Chemical parameter	Unit	$U_{a\text{-lab}}$	$U_{a\text{-repl}}$	$U_{s\text{-repl}}$		$U_{m\text{-repl}}$			Rel meas var			Mean	
		Expanded relative uncertainty from analysis (incl. bias) (95%) (from laboratory QA/QC program) (%)	Expanded relative uncertainty from analysis (95%) COV of laboratory replicates (%)	Expanded relative uncertainty from sampling (95%) RANOVA of replicates (%)		Expanded relative uncertainty from measurement (95%) (sampling + analysis) COV of replicates (%)			Relative measurement variability (sampling + analysis) COV of replicates (%)			Mean value/concentration during experiment	
			KVA	SVA	KVA	SVA	KVA	KVA_{AS}	SVA	KVA	KVA_{AS}	SVA	KVA
Alkalinity	mmol/l	10	1.2	<0	2.3	–	2.6	0.0	–	1.3	0.0	–	0.4
Conductivity	mS/m	5	0.4	0.9	0.3	1.0	0.5	0.7	0.5	0.2	0.3	9.9	292
P_{tot}	mg/l	25	24.7	<0	<0	–	9.8	15.6	–	4.9	7.8	–	0.027
Nitrate	mg/l	15	–	–	–	3.1	2.9	4.0	1.5	1.4	2.0	0.13	23.04
Sulphate	mg/l	15	–	–	–	7.3	4.7	7.3	3.7	2.3	3.6	17.2	1217
Copper	µg/l	≈20	–	–	–	29.5	–	–	14.7	–	–	0.39	–
Vanadium	µg/l	≈20	–	–	–	23.1	–	–	11.5	–	–	0.08	–
Zink	µg/l	≈20	–	–	–	28.5	–	–	14.2	–	–	4.26	–
Uranium	µg/l	≈20	–	–	–	9.2	–	–	4.6	–	–	0.44	–

KVA_{AS} represent the results from the replication experiment using the Aloha Sampler™ at KVA sampling target fourfold increase in duration allows larger temporal variability in stream flux to influence the sampling, which perhaps could be another reason for the slight increase in relative measurement uncertainty when using the Aloha Sampler™ for primary sample extraction.

Discussion

The results from this study explicitly showed how important experimental design is for proper evaluation of sampling system variabilities. The design and adherence to guidelines and norms will directly influence the numerical evaluation of the results and therefore need to be fully transparent and well defined. The implication of the results from the two separate duplicate sampling experiments indicate that the current water sampling and analytical process is showing acceptable variability under repeatability conditions, or as described by a strictly defined sampling protocol. However, the temporal spacing between primary samples at KVA led to high measurement system variability, due to the large temporal variabilities in the stream water. This study shows that replication experiments together with duplicate sampling under repeatability conditions can result in lower sampling and measurement variability estimations than duplicate experiments, taking the ambiguity in the sampling protocol in to account. The main reason for this was the large temporal changes at the sampling locations, stemming from variability in water quality.

The replication experiment using the Aloha Sampler™ showed higher sampling variability than the current grab sampling method for all parameters except alkalinity. This is likely because the Aloha Sampler™ covers a larger part of the cross-stream heterogeneity and collects increments over longer periods of time. The reason that this does not increase the measurement variability for alkalinity is likely due to the lower overall variability for this parameter, compared to all of the other analytes. The spatial variability for alkalinity might be larger than the temporal variability, leading to larger variabilities for the grab sampling approach as applied to one more-or-less randomly selected locations in the stream, than for the Aloha Sampler™ (collecting increments from the complete width of the stream). One advantage of using the Aloha Sampler™ for routine sampling is a wider coverage of the stream width, as well as covering longer time intervals and thereby temporal concentration changes. With the existing diurnal and seasonal changes in analyte concentrations and outflows from the clarification pond, this small extension in time coverage would likely still lead to large sampling variabilities when the temporal ambiguity in the current sampling protocol is considered.

It is not only the practical sample extraction method that needs to be evaluated to determine the complete measurement variability. The environmental sampling protocols, which dictate when and how often samples shall be collected, are just as important. This study gives a clear indication that the temporal variabilities in natural stream flow, outflow from the industrial operation, and analyte concentrations need to be considered when designing improved sampling protocols, both with regard to sampling interval and sample extraction method. It is essential to ensure that the variability in 'true' concentrations are intercepted and represented correctly. For sampling targets with large variabilities in concentration and stream flow, or variabilities in overflow from, for example, a clarification pond, the sampling protocol must be adapted to these empirical conditions. The sampling and measurement variability for especially KVA could be improved by a more concise sampling protocol that counteracts temporal variabilities. One recommended solution would be to install an automated sampler collecting increments from the stream, either at regular time intervals or based on volume-proportional composite sampling. An important learning experience from TOS, also revealed in this study, is that standards and norm-giving manuals are not necessarily a guarantee that recommended sampling approaches are correct or representative. The evaluation of total measurement variability through well designed experiments is therefore always necessary.

To allow for continuous control of the current sampling systems for recipient water sampling, systematic duplication or replication experiments at regular time intervals is recommended. However, the ambiguity in existing sampling protocols, as described by Esbensen and Wagner (2016), Lyn et al. (2007), and Ramsey (1998), must be considered to allow valid estimations of the complete sampling and measurement variability. In comparison, to follow the duplicate method as described by Grøn et al. (2007), strictly (extracting duplicates specifically under repeatable conditions) will unavoidably underestimate the complete sampling variability derived by most sampling protocols.

Conclusions

This study shows that the specific approach for estimating sampling variability will affect the results significantly. The scope of each specific investigation (experimental design, time duration, sampling frequency) must be carefully defined so that the most appropriate experimental method that will best match the pre-set objectives can be selected. This study concluded that replication experiments together with duplicate sampling under repeatability conditions resulted in lower sampling and measurement variability estimations than duplicate experiments, taking the ambiguity in the sampling protocol into account. Hence, to correctly evaluate all of the variabilities affecting the measurement system, all ambiguities in the sampling protocol need to be included in the experimental design. The time-varying nature of stream water makes sampling a complex undertaking, and indicates that the existing official protocols should be revised.

The results indicate that the relative sampling-and-measurement variability is acceptable under repeatability conditions, both using the current sample extraction method and using the Aloha Sampler™. However, the duplicate experiment at KVA, where the larger ambiguity in the sampling protocol was taken in to account, indicated that the measurement variability for several analytes were above the 20% threshold. This shows that even if the sample extraction can produce repeatable results over short times periods, ill-reflected ambiguity in the current sampling protocol for KVA is likely leading to adversely large measurement variabilities for some analytes. The results of this study do not enable definite conclusions regarding the reliability of the Aloha Sampler™ compared to traditional grab sampling. The Aloha Sampler™ replication experiments shows larger sampling variability (for most analytes) than the current sample extraction method, but this likely reflects reality better.

A recommended solution for counteracting large temporal variabilities is to install automated samplers, regularly collecting increments from the stream and combining them into composite samples. With a composite sampling method and regular analysis, the coverage of temporal changes in stream flow and analyte concentration would be significantly improved, compared to any manual sampling approach. Both duplicate and/or replication experiments (where samples are spaced relative to the ambiguity in the sampling protocol) are practical and worthwhile methods for initial, as well as regular, quality assessment of recipient water sampling. To allow for a complete evaluation of sampling and analysis variability, representative for all seasonal variations in flow and analyte concentrations, the experiments should be applied continuously and regularly over an entire year.

Acknowledgements The authors acknowledge the environmental sampling staff and laboratory staff at LKAB for all their help during the experiments within this study. We also thank LKAB for allowing publication of this study.

References

Botta F, Lepot B, Leoz-Garziandia E, Morin A (2012) Estimation of sampling uncertainty in lake-water monitoring in a collaborative field trial. Trends Anal Chem 36:176–184

Danish S (2013) DS 3077 representative sampling—horizontal standard. Danish Standards Foundation, Nordhavn

Esbensen KH (2016) Pierre Gy (1924–2015): a monumental scientific life. TOS Forum 6:39–46

Esbensen KH, Paasch-Mortensen P (2010) Process Sampling: theory of sampling—the missing link in process analytical technologies (PAT). In: Bakeev KA (ed) Process analytical technology, 2nd edn. Wiley, Chichester, pp 37–80

Esbensen KH, Wagner C (2014) Theory of sampling (TOS) vs. measurement uncertainty (MU)—a call for integration. Trends Anal Chem 57:93–106. https://doi.org/10.1016/j.trac.2014.02.007

Esbensen KH, Wagner C (2016) Sampling quality assessment: the replication experiment. Spectr Eur Asia 28(1):20–25

Esbensen KH, Friis-Petersen HH, Petersen L, Holm-Nielsen JB, Mortensen PP (2007) Representative process sampling—in practice: variographic analysis and estimation of total sampling errors (TSE). Chemometr Intell Lab Syst 88:41–59. https://doi.org/10.1016/j.chemolab.2006.09.011

Gammons CH, Milodragovich L, Belanger-Woods J (2007) Influence of diurnal cycles on metal concentrations and loads in streams draining abandoned mine lands: an example from High Ore Creek, Montana. Environ Geol 53:611–622. https://doi.org/10.1007/s00254-007-0676-z

Grøn C, Bjerre Hansen J, Magnusson B, Nordbotten A, Krysell M, Jebjerg Andersen K, Lund U (2007) Uncertainty from sampling - a nordtest handbook for sampling planners on sampling quality assurance and uncertainty estimation. Nordtest TR-604. Nordic Innovation Centre, Oslo. ISSN 0283-7234

Guigues N, Desenfant M, Lalere B, Vaslin-Reimann S, Eyl D, Mansuit P, Hance E (2016) Estimating sampling and analysis uncertainties to assess the fitness for purpose of a water quality monitoring network. Accredit Qual Assur 21:101–112. https://doi.org/10.1007/s00769-015-1186-4

Gy P (1979) Sampling of particulate materials, theory and practice. Elsevier, Amsterdam. ISBN: 9780444601353

Gy P (1998) Sampling for analytical purposes. Wiley, Chichester

Gy P (2004) Sampling of discrete materials—a new introduction to the theory of sampling I. qualitative approach. Chemometr Intell Lab 74:7–24

Henjum MB, Hozalski RM, Wennen CR, Novak PJ, Arnold WA (2010) A comparison of total maximum daily load (TMDL) calculations in urban streams using near real-time and periodic sampling data. J Environ Monit 12:234–241

Lyn JA, Ramsey MH, Dmanat AP, Wood R (2007) Empirical versus modelling approaches to the estimation of measurement uncertainty caused by primary sampling. Analyst 132(12):1231–1237

Nimick DA, Cleasby TE, McCleskey RB (2005) Seasonality of diel cycles of dissolved trace-metal concentrations in a Rocky Mountain stream. Environ Geol 47:603–614

Pitard FF (1993) Pierre Gy's sampling theory and sampling practice: heterogeneity, sampling correctness, and statistical process control. CRC Press LLC, Boca Raton (**ISBN : 0-8493-8917-8)**

Ramsey MH (1998) Sampling as a source of measurement uncertainty: techniques for quantification and comparison with analytical sources. J Anal Atom Spectrom 13:97–104

Ramsey C (2014) The Aloha Sampler™: concept, objective, design and implementation. TOS Forum 3:12–14

Ramsey MH, Thompson M (2007) Uncertainty from sampling, in the context of fitness for purpose. Accred Qual Assur 12:503–513. https://doi.org/10.1007/s00769-007-0279-0

Ramsey MH, Thompson M, Hale M (1992) Objective evaluation of precision requirements for geochemical analysis using robust analysis of variance. J Geochem Explor 44:23–36

Ramsey MH, Geelhoed B, Wood R, Damant AP (2011) Improved evaluation of measurement uncertainty from sampling by inclusion of between-sampler bias using sampling proficiency testing. Analyst 136(7):1313–1321

Rode M, Suhr U (2007) Uncertainties in selected river water quality data. Hydrol Earth Syst Sc 11:863–874

Westerberg I, Huseby Karlsen R (2017) Vattenföringsmätningar i Liukattijoki och Mertaseno. Yearly technical report for LKAB. Swedish Environmental Research Institute, Stockholm

Ylipää M, Björnfot M (2018) Koncentrationsvariation i bräddvattnet vid utökad magasinering i klarningsmagasinet. Internal technical report. LKAB, Kiruna