Fisheries and Oceans Canada | Pêches et Océans Canada
Ecosystems and Oceans Science | Sciences des écosystèmes et des océans

Canadian Science Advisory Secretariat (CSAS)

Research Document 2025/017

National Capital Region

Considerations of Bias and Precision Associated with ISE Based Measurements of Sediment 'Free' Sulfide

F.H. Page and D.K.H. Wong

Fisheries and Oceans Canada
Biological Station
125 Marine Science Dr.
St. Andrews, N.B., E5B 0E4

February 2026

Canada

Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
DFO.CSAS-SCAS.MPO@dfo-mpo.gc.ca

ISSN 1919-5044
ISBN 978-0-660-75878-7 Cat. No. Fs70-5/2025-017E-PDF

Correct citation for this publication:

Page, F.H., and Wong, D.K.H. 2026. Considerations of Bias and Precision Associated with ISE Based Measurements of Sediment 'Free' Sulfide. DFO Can. Sci. Advis. Sec. Res. Doc. 2025/017. v + 82 p.

Aussi disponible en français :

Page, F.H. et Wong, D.K.H. 2026. Considérations relatives au biais et à la précision associés aux mesures par EIS des sulfures « libres » dans les sédiments. Secr. can. des avis. sci. du MPO. Doc. de rech. 2025/017. v + 85 p.

TABLE OF CONTENTS

ABSTRACT

This document utilizes literature and new data to examine the bias (systematic error), precision (non-systematic or random error; indicated by standard deviations and coefficients of variation (CV)) and total uncertainty or accuracy (includes bias and precision) associated with aspects of Ion Selective Electrode (ISE) measurements as applied to the determination of “free” sulfide in marine surficial sediment associated with aquaculture operations. The examination concludes that a careful and experienced analyst should be able to routinely achieve ISE sulfide measurements that have a total uncertainty of 10% or less when standard solutions, electrodes and samples are not stored, standard solutions are titrated and millivolt readings are made within 2 minutes of electrode insertion into a sample and electrodes are regularly checked and calibrated (at least once every 3 h). The examination also concludes that storage of standard solutions, storage of calibrated electrodes, storage of sediment samples and failing to account for sediment water content can all increase this uncertainty quite significantly, by increasing bias and reducing precision (increasing CVs). Bias is typically less than 10% and often about 5% when storage times are near zero, the bias is routinely greater than 10% when storage times exceed 4 h and can often exceed 20%. The actual magnitude and direction of bias varies between samples and storage times, and is not presently known or predictable, and as a result measurements on stored samples cannot be reliably adjusted for bias after the fact. The precision of ISE sulfide measurements is largely independent of sample storage time with values of the CV ranging from <5% to >30%. The precision associated with well conducted analyses is much greater (lower CVs) than the CVs obtained from multiple measurements from spatially separated in-situ sediment samples. Although highly variable, the CVs associated with spatially separated in situ measurements generally increases with the magnitude of the separation distance between samples. Variation on the scale of a fish farm can be of order 100% whereas the variation associated with a well conducted analysis that is not influenced by storage factors is of order 10%. The high CVs of the spatially separated measurements of in situ sediments is assumed to be generated by inherent natural variability in sediment characteristics on the scales of meters (within station variability) to hundreds of meters (within farm variability). The implication to environmental management is that well defined standard operating procedures need to be adopted that ideally do not include storage of standards, electrodes or samples and that environmental monitoring sample designs and summary statistics for sediment pore water sulfide need to robustly take into consideration the inherent natural variability of the environment so monitoring results are reproducible and hence credible.

INTRODUCTION

One of the objectives of Canadian Federal and Provincial environmental regulators of marine aquaculture in Canada is to help manage and limit the environmental impact of organic loading by marine finfish net pen and suspended shellfish culture activities. A main regulatory goal is to limit the degree of loss of infauna biodiversity within the surficial sediments exposed by aquaculture induced organic loading.

Obtaining measurements of biodiversity has long been known to require specialized knowledge of taxonomy and to be time-consuming and expensive; hence a cheaper and quicker indicator of biodiversity reductions was desired (Wildish et al., 2004a). The science supporting the development of the aquaculture environmental regulations recognized that in soft sediments there is a gradient in biodiversity with high biodiversity corresponding to well oxygenated sediments and low biodiversity with poorly oxygenated sediments. The science also recognized that measurement of sediment oxygen concentrations was challenging, particularly with the technology available at the time (other methods are now being explored), and that there was a relationship between sediment pore-water sulfide concentration and biodiversity. The relationship between biodiversity and ISE sulfide concentration being, particularly when combined with sediment redox (Eh; oxidation – reduction) measurements, one of the better indicators of soft-sediment infauna biodiversity (Hargrave et al., 1997).

Fisheries and Oceans Canada and several Provinces have adopted sulfide as the regulatory indicator of benthic infauna biodiversity impact associated with organic discharges from finfish farms (Aquaculture Activity Regulations (AARs), provinces). The regulatory questions asked and the sampling design vary between jurisdictions. However, in all jurisdictions samples of sediment are collected, processed, and analyzed for total free pore water sulfide using the ISE methodology introduced by Wildish et al. (1999) and in all jurisdictions the measurements are combined in some way to calculate an index of impact. For example, in New Brunswick (NB) and Nova Scotia (NS) the data are combined to calculate a mean concentration of sulfide that is thought to be biased toward the highest impact areas (i.e., samples are taken adjacent to net pens with the largest number of fish). In all cases the calculated indices have an uncertainty or error associated with them that results from the propagation of the measurement errors through the estimation process (i.e., through the mathematical relationships underlying the estimations).

In all Canadian jurisdictions, the regulatory practice is to compare ISE sulfide measurements to predefined sulfide concentration thresholds that represent a progressive decrease in sediment oxygen content and infauna biodiversity to the completely unacceptable anoxic level of 6,000 µM sulfide. The various sulfide threshold concentrations suggested by science and adopted by Canadian regulators are summarized in Appendix 3. Sampling designs have been specified and when sulfide concentrations exceed the established thresholds, specific regulatory actions are triggered (e.g., Table A-NB2).

As with any measurement and index, knowledge of the associated uncertainty is essential to making meaningful interpretations of the values. Without this knowledge poor decisions can be made (Hall and White, 2020). For example, it could be concluded a site mean sulfide concentration exceeds a specific threshold when in fact it does not. Ideally the sulfide measurements, along with the indices and interpretations derived from them, are highly reproducible. This includes taking and analyzing measurements using collection and analyses procedures in ways that have sufficiently small and known influences and uncertainties that allow exceedances of specified thresholds to be robustly detected. The authors of the original documents proposing the ISE method were aware of many of the factors that could potentially cause measurement errors and made efforts to explore and quantify some of these (Wildish

et al., 1999; 2004a). Once the method became adopted as a recognized regulatory measure and became more widely used, additional experience and perspective was gained concerning the measurement uncertainties and the factors contributing to them. Use of the ISE sulfide indicator as a regulatory measure needs to consider that the measurement technique must be meticulously described and followed as well as take into consideration the natural magnitude of temporal and spatial variation in the measurements, especially in relation to the spatial and temporal variability inherent in the sulfide and biodiversity measurements used to establish the thresholds.

The data points defining the existing relationship between ISE sediment sulfide and biodiversity has considerable scatter, with the scatter being related to several factors including measurement error and natural variability. Ideally, the scatter can be reduced by being defined using data collected using appropriate sample designs and measurement techniques. The degree of reduction needed depends on the accuracy required by regulators and the ability of sample collectors and analysts to achieve unbiased and high precision in their ISE measurements.

The purpose of the following analyses is to update and more fully explore the uncertainty associated with sulfide measurements made using the ISE sulfide analysis method, place this in the context of the variability in ISE measurements obtained from natural situations, and comment on the implications of these considerations to the ability to detect exceedances of specified thresholds. This document is one of three documents prepared in response to a request for science advice in relation to ISE determinations of sediment pore water free sulfide concentration. The terms of reference asked the following questions:

1. What are the effects of sediment sample storage time and conditions on the measurement of total free sulfide as compared to total free sulfide measured immediately upon sample collection?
2. Are these relationships consistent across sediment types and/or total free sulfide concentrations?
3. Is there a combination of storage conditions and storage times post collection that would result in expected total free sulfide measurements within ±5%, 10% and 15% of the value obtained from measuring total free sulfide immediately following sediment sample collection?
4. Are there steps in the ion-selective electrode (ISE) total free sulfide measurement protocol that are open to interpretation by the analyst and to which differences will result in different measured concentrations of total free sulfides?
5. Review ISE total free sulfide measurement methodologies and develop standard procedures for sample storage time, storage conditions, and analyses.
6. In the consideration of the above questions, characterize the method variability in the context of natural in situ, spatial variability of sediment sulfide levels.

This document focuses on question 6 and contributes to questions 1, 2, and 3. Details of the ISE method is the focus of Wong and Page (2025a) and the storage time on sediment sulfide concentration is the primary focus of Wong and Page (2025b).

SECTION 1: UNCERTAINTY OVERVIEW

From a general perspective, we usually never know the true value of what is being measured; we can only estimate this value through measurements and all estimates and measurements

have uncertainty associated with them (Taylor, 1997; Berendsen, 2011). Put another way, all measurements are estimates of a true value, with the measured value equaling an unknown true value plus an uncertainty (sometimes referred to as error) (Eq. 1) and all measurements should be reported as the measured value plus or minus an uncertainty. The specification of error or uncertainty associated with the measurements should include an indication of the probability distribution associated with the uncertainty (Berendsen, 2011). Measurements without associated quantification of its uncertainties have limited use because there is no way of knowing how true or reliable the measurement is (i.e., how close the estimate is to the true value) (Taylor, 1997; Berendsen, 2011).

Knowledge of the uncertainties associated with measurements is particularly relevant to regulatory decision making. Quantification of the errors and uncertainties and inclusion of this information in the decision making contributes to the credibility of the decision since differences between measurements can be interpreted as being either within or outside the expected limits of the measurement uncertainty.

Uncertainty is often considered to consist of two main components: systematic error or bias and non-systematic error or precision (Taylor, 1997). Systematic error (bias) refers to a consistent difference between the true value and the measured or estimated value. Non-systematic error is the random variation associated with a measurement. This error has a specific frequency distribution and the parameters that define the distribution need to be established. The total error or uncertainty includes both the bias and precision and is referred to as the measurement accuracy. A more detailed and mathematical consideration of these concepts is given in Appendix 2.

From the specific perspective of sulfide measurements, ISE determinations of sediment pore water ‘free’ sulfide also have errors and uncertainties associated with them. The uncertainties in ISE measurements are generated by multiple factors with some factors resulting in biased measurements and other factors contributing to the precision of the measurements. Many of these factors have been identified and discussed in Wong and Page (2025a, b) as well as in earlier publications (Chang et al., 2014; Cranford et al., 2020; Hargrave et al., 2008b). Some of the factors that have been considered are associated with laboratory procedures, such as probe calibration, standard solution preparation, probe storage, and sample storage. Other factors are related to sample collection procedures such as grabs or cores and natural temporal and spatial inhomogeneities in the environment (Cranford, et al., 2011).

Ideally, the sulfide measurements should be as accurate as possible i.e., be true and consistent with little or no bias and high and repeatable precision (low random error). In practice the measurements need to be sufficiently accurate i.e., the accuracy is sufficient to robustly support the inferences made from the measurements. In the case of Canadian Federal and Provincial aquaculture regulation the accuracy should be sufficient to enable robust classification of farm sites based on specified thresholds (Appendix 3).

The purpose of the analyses presented here is to explore the magnitude of some of the sources of uncertainty and to contrast methodological uncertainty to environmental variability. Details of the ISE sediment sulfide analyses are reviewed in (Wong and Page, 2025a,b). The approach taken has been to begin with a consideration of the accuracy of individual ISE measurements made under specific controlled conditions and progress toward uncertainties associated with more operational measurements.

SECTION 2: BASE UNCERTAINTY OF ISE MEASUREMENTS

FIRST IMPRESSIONS

It is generally recognized that the ISE sulfide measurement process is not the most accurate method for measuring pore water sulfide (Hargrave et al., 2008b; Schaaning and Hansen, 2005). Spectrophotometric and iodometric titration methods are sometimes considered to be more accurate (Schaaning & Hansen, 2005).

The ISE method has been characterized by some as an order of magnitude method (Rundle 2011). Although this conveys the message that the method is considered to produce only rough estimates of sulfide concentration, the precise meaning of "order of magnitude" was not defined. In mathematical terms the order of magnitude of a number is the number rounded to the nearest power of ten, i.e., if the number is between 10 and 99 the order of magnitude is 10, if the number is between 100 and 999 the order of magnitude is 100, and so on. In this context 75 µM is of order 10 µM, 760 µM is of order 100 µM, 1,500 µM is or order 1,000 µM and 5,100 µM is also of order 1,000 µM. This is the meaning typically used in science and in this document. In a project management context, initial scoping estimates are often referred to as rough order of magnitude estimates. In this context some consider this means the measurements are accurate to within ±50%, i.e., a measurement of 1,000 µM could be somewhere between 500 (-50%) and 1,500 (+50%) µM and to others it means the uncertainty is ±35% (Rowe, 2015). To others the order of magnitude terminology is defined to be within -25% and +75% of the initial estimate (Maley, 2012). For example, if the sulfide measurement value is 100 µM the true value is assumed to be between 25 µM and 175 µM or if the measurement value is 4,000 µM the true value is assumed to be between 1,000 µM and 7,000 µM. No matter the definition, an order of magnitude measurement is not ideal for a regulatory regime that desires to robustly detect sulfide values above threshold values that are within the same order of magnitude, particularly when the regulatory consequence may be no restocking of a farm.

Ideally one can obtain a perspective on the degree of uncertainty associated with a measurement from the number of significant digits associated with the reported measurement. This is only a reliable estimate if the source follows proper significant digit reporting protocols. Table 1 gives example calculations of accuracy for measurements reported to levels of 0.1, 1, 10 and 100 µM. In practice, many reports of ISE sulfide measurements report sulfide concentrations to the nearest 1 µM and this implies an uncertainty of ±1 µM. Assuming the uncertainty or accuracy is unbiased and the variation or errors are normally distributed with a standard deviation of 1 µM, then 95.4% of the time the true sulfide concentration lies between the recorded concentration ± 1 µM (Table 1). This implies the relative accuracy of the measurement, the precent uncertainty, decreases with concentration, i.e., an error of ±1 µM associated with a measurement of 100 µM represents a 1% uncertainty, whereas an error of ±1 µM associated with measurements of 1,000 and 10,000 µM represent relative uncertainties of 0.1% and 0.01% respectively. When ISE sulfide measurements are reported to the nearest 10 µM the implication is that an uncertainty of ±10 µM, i.e., the 95.4% of the time the true sulfide concentration lies between the recorded concentration minus 10 µM and the recorded measurement plus 10 µM. This implies a relative accuracy of 10% for a measurement of 100 µM and accuracies of 1% and 0.1% for measurements of 1,000 and 10,000 µM respectively.

As discussed later in this document, estimates derived in this manner are probably not reliable for most reported ISE measurements because reported measurement values often do not follow the significant digit rules are often not intended to convey measurement accuracy. For example, an uncertainty of ±1 µM is probably an underestimate of the uncertainty.

Table 1: Summary of uncertainties implied by the number of significant digits reported for a range of sulfide concentration measurements. The accuracy is assumed to represent one standard deviation in the unbiased error.

Reporting Accuracy of Sulfide Conc. (mM)	Sulfide Concentration			Implied Absolute Uncertainty (µM)	Implied Relative Uncertainty (%)
	True Sulfide (µM)	Lower Limit (µM)	Upper Limit (µM)		
0.1	100	99.9	100.1	0.1	0.100
	500	499.9	500.1	0.1	0.020
	1000	999.9	1000.1	0.1	0.010
	5000	4999.9	5000.1	0.1	0.002
	10000	9999.9	10000	0.1	0.001
1	100	99	101	1.0	1.000
	500	499	501	1.0	0.200
	1000	999	1001	1.0	0.100
	5000	4999	5001	1.0	0.020
	10000	9999	10001	1.0	0.010
10	100	90	110	10.0	10.000
	500	490	510	10.0	2.000
	1000	990	1010	10.0	1.000
	5000	4990	5010	10.0	0.200
	10000	9990	10010	10.0	0.100
100	100	0	200	100.0	100.000
	500	400	600	100.0	20.000
	1000	900	1100	100.0	10.000
	5000	4900	5100	100.0	2.000
	10000	9900	10100	100.0	1.000

Fortunately, it is recognized that with appropriate attention to analytical detail and quality control, it may be possible to achieve uncertainties to about 8% or less (Rundle 2011). Some have claimed the ISE method can achieve absolute accuracies of ±3 µM (Hargrave et al., 1993b). Others have indicated the reproducibility of another method of sulfide measurement to be between 2.3% and 6.2% (Schaaning and Hansen, 2005) based on median pS values (3.2 ro 8.7) of sediment sulfide in Norwegian waters and assuming values were within about 0.2 pS units (pS = - log[$\sum$S]) of measured values.

In the following sections we explore the emprical evidence related to the accuracy of ISE sulfide measurements that have been made in association with various measuring activities made in relation to Canadian aquaculture activities.

THEORETICAL CONSIDERATIONS ASSOCIATED WITH THE CALIBRATION RELATIONSHIP

The foundation of the ISE method is the Nernst Equation (Eq. 1) and the calibration relationship based on this equation. In the Nernst equation V_{eq} is the

$$V_{eq} = \frac{R \cdot T_K}{z \cdot F} \cdot Ln\left(\frac{C_{out}}{C_{in}}\right) \quad \text{Eq. 1}$$

equilibrium potential, also known as the Nernst potential for a specified ion, R is the universal gas constant equal to 8.314 J·K-1·mol-1 (Joules per Kelvin degree per mole of ion), T_K is the temperature of the solution containing sulfide in degrees Kelvin ($T_K = T_C + 273.15$ where T_C is the temperature in degrees Celsius), z is the valence or charge of the ion and has a value of -2 for sulfide (S-2), F is the Faraday constant equal to 96485 C·mol-1, C_{out} is proportional to the concentration of the ion on the outside of the electrode membrane and C_{in} is proportional to the concentration of the ion on the inside of the electrode membrane. Note: C_{out} and C_{in} are actually the electrical activities associated with the ion and these are assumed to be proportional to the ion concentrations.

The calibration relationship is based on the expanded and linearized Nernst Equation (Eq. 2).

$$\begin{aligned} V_{eq} = \frac{R \cdot T_K}{z \cdot F} \cdot Ln\left(\frac{C_{out}}{C_{in}}\right) &= \frac{R \cdot T_K}{z \cdot F}\left[Ln(C_{out}) - Ln(C_{in})\right] \\ &= -\frac{R \cdot T_K}{z \cdot F} \cdot Ln(C_{in}) + \frac{R \cdot T_K}{z \cdot F} \cdot Ln(C_{out}) \end{aligned} \quad \text{Eq. 2}$$

Replacing $-\frac{R \cdot T_K}{z \cdot F} \cdot Ln(C_{in})$ with β_0 and $\frac{R \cdot T_K}{z \cdot F}$ with β_1 gives the linear calibration relationship (Eq. 3). This can be used by using least squares regression to estimate the parameters or be approximated by a series of linear segments.

$$V_{eq} = \beta_0 + \beta_1 \cdot Ln(C_{out}) \quad \text{Eq. 3}$$

When the natural logarithm (Ln) is replaced with the base ten logarithm (Log), equation 3 becomes equation 4 with the parameters now being redefined as $\beta_0 = -\frac{R \cdot T_K}{z \cdot F} \cdot 2.303 \cdot Log(C_{in})$ and $\beta_1 = 2.303\frac{R \cdot T_K}{z \cdot F}$ (Rundle, 2011).

$$V_{eq} = \beta_0 + \beta_1 \cdot Log(C_{out}) \quad \text{Eq. 4}$$

Rearranging equations 1, 3 and 4 and converting the rearranged equations into their exponential forms to solve for the concentration (C_{out}) gives the concentration of the ion as (Eq. 5). These concentrations are technically ionic activities which in the case of very dilute solutions are effectively concentrations of the ion of interest (Rundle 2011). In this document the term concentration rather than activity or effective concentration is used since the two values are almost identical for the concentrations involved in ISE measurements (Wong and Page, 2025a) and the term concentration is used in most aquaculture literature referring to ISE sulfide measurements.

$$C_{out} = C_{in} \cdot e^{V_{eq} \cdot z \cdot F / R \cdot T_K} \quad \text{Eq. 5}$$

$$C_{out} = e^{(V_{eq} - \beta_0)/\beta_1}$$

$$C_{out} = 10^{(V_{eq} - \beta_0)/\beta_1}$$

The Calibration Relationship

The foundation of the ISE sulfide method described by Wildish et al. (1999) is the calibration of the ion selective probe and calibration relationship (Eq. 6).

The calibration relates a millivolt reading (V_S) to the logarithm of a nominal concentration or "known" concentration of sulfide (S_{NC}), i.e., solutions whose sulfide concentration is constructed by dilution of a standard solution of sulfide whose concentration has been measured by a titration and has an error or uncertainty (ε) associated with it.

Figure 1 shows an example of a calibration relationship with the left panel showing the millivolt reading verses the logarithm of the standard sulfide concentrations and the right panel showing the millivolt reading verses the untransformed sulfide concentration.

$$V_S = \beta_0 + \beta_1 log_{10}(S_{NC}) + \varepsilon \quad \text{Eq. 6}$$

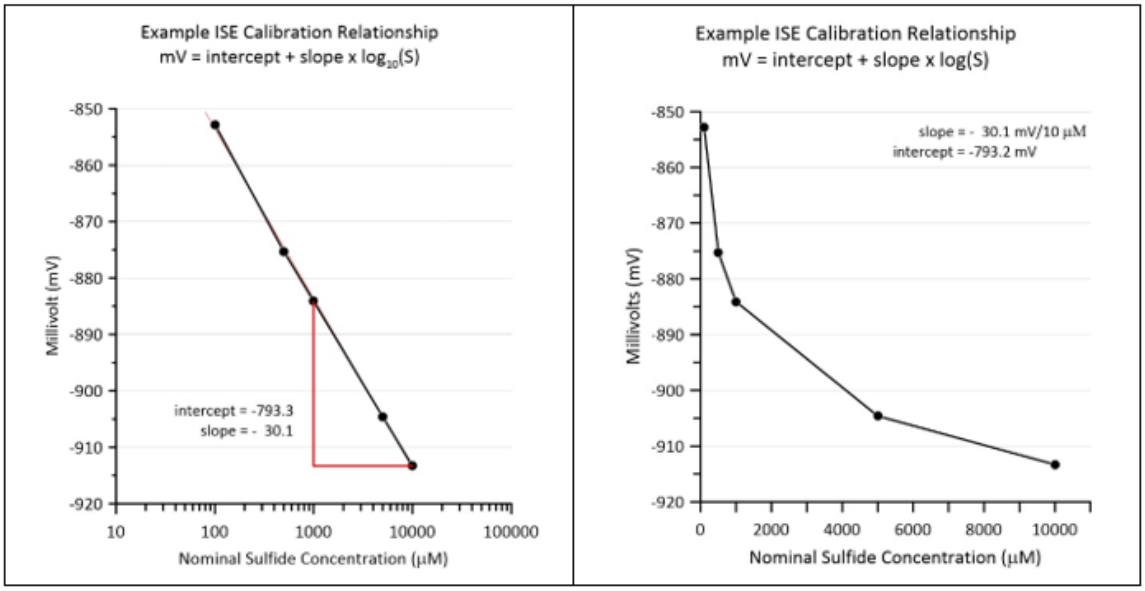

Figure 1: Example of an ISE calibration relationship plotted with the dependent variable (millivolts) against the dependent variable sulfide concentration. Left panel is log₁₀(sulfide concentration). Right pane is sulfide.

The calibration relationship is only valid for the sulfide concentrations for which the Nernst Equation applies, for the range of standard concentrations used in the calibration relationship and for the portion of the log-linear calibration relationship that is linear with a slope within the probe manufacturers acceptable range. The log linear Nernst relationship breaks down at very

low and high concentrations of ions. For the ThermoScientific Probe, the lower concentration limit is 1 x 10^{-5} M (i.e., 10 µM; TFS 2007). The upper limit is not stated. The theoretical slope varies with the temperature of the standard solutions but at typical room temperatures is about 29.

A measurement of sulfide in a sample, is obtained by taking a millivolt reading from the sample and the estimate of the concentration of sulfide is obtained by using . 7, which is the inverse of the calibration equation (Eq. 6).

$$S_M = 10^{(mV \pm \varepsilon - \beta_0)/\beta_1} \qquad \text{Eq. 7}$$

There are two approaches commonly implemented in association based on equation 7. One is the segmented approach and the other is the regression approach.

Segmented Approach: The voltmeters used in ISE sulfide measurements typically do not conduct a regression through the standard calibration points; instead, the meters treat the calibration points as a series of adjoining two-point linear segments, with a linear relationship between mV and $\log_{10}$ sulfide derived for each sequential pair of coordinates. The segment bracketing the mV reading for a sample is used to estimate the sample sulfide concentration. Ideally the slope of each segment is the same, but errors in millivolt readings usually result in the slopes being marginally different. This introduces some error into the sulfide measurements made on samples, but because only two calibration points are used in the definition of each segment a statistical error in the slope and inverse sulfide estimates cannot be made other than through the approach outlined above.

Regression Approach: The regression approach estimates the values of β_0 and β_1 by conducting a linear regression through the calibration points. This ensures the slope is constant throughout the range of standard concentrations used, allows an error to be estimated for each parameter and therefore enables the possibility of estimating an error for each sulfide concentration derived from the regression equation. The error in the sulfide concentration is not constant over the calibration range due to the non-linearity of the underlying relationship and due to the assumptions associated with inverse linear regression prediction. Details of the inverse regression approach have not been explored here, in part because many analysts do not seem to use it, and because there are other factors that generate larger errors than the difference between regression and segmented estimates.

In both approaches the relationships should be examined for deviations from a linear relationship. The log linear Nernst relationship is known to break down at very low and high concentrations of ions, such that the linear relationship between mV and ion concentration no longer holds at ion concentrations below and above specific ion concentrations. As an example, the ThermoScientific Probe manual indicates the lower limit for their sulfide electrodes is 1 x 10^{-5} M (i.e., 10 µM; TFS 2007). The upper limit is not stated.

As indicated above the calibration relationship is based on measurements that contain measurement errors and therefore the estimates of sulfide concentration have an uncertainty associated with them. Rundle (2011) indicated that for divalent ions such as sulfide, the magnitude of error is expected to be about 8% per 1 mV in properly calibrated electrodes. In practice, the actual error depends on the probe, voltage meter, and the analyst's implementation of the various calibration and measurement procedures.

The error associated with a measured voltage propagates through the calibration relationship resulting in an error associated with the derived sulfide (Berendsen, 2011). In the following paragraphs we examine the propagation of the measurement errors through the calibration

relationship. The non-linear nature of the ISE calibration relationship amplifies the initial uncertainties.

Uncertainty Associated with Millivolt Reading Error

The effect of error in millivolt readings on sulfide estimates can be explored by keeping the calibration coefficients constant and varying the millivolt value. Absolute and relative errors in sulfide concentration when errors in the calibration coefficients are assumed to be zero (e.g., segment calibrations) are given in Appendix 6. The upper (S_M^u) and lower (S_M^l) values bracketing a sulfide measurement are given by Eq. 8 and when the relative uncertainty is defined as the sulfide measurement with error divided by the measurement without error, the uncertainty is given as $U_{rel} = 10^{\pm\varepsilon/\beta_1}$.

$$S_M^u = S_M^{\varepsilon=0} 10^{\varepsilon/\beta_1} \, , \, S_M^l = S_M^{\varepsilon=0} 10^{-\varepsilon/\beta_1} \qquad \text{Eq. 8}$$

Appendix 6 contains calculations of the effect of a range of millivolt uncertainties on log sulfide and sulfide estimates. The calculations express the effect in terms of bias and precision. For no millivolt error, bias is zero and there is no variation i.e., precision is perfect. For millivolt errors of ± 0.1 mV the corresponding $\log_{10}$(S) values have an absolute difference of 0.003 log units from the true value and a relative difference of less than 0.17%. For millivolt errors of ±1 mV the corresponding log10(S) values have an absolute difference of 0.034 log units and a relative difference of less than 1.7%. The table also illustrates that the absolute differences are independent of sulfide concentration and that the relative differences decrease with increasing concentration for a given millivolt uncertainty. In all cases the absolute and relative bias is zero.

The absolute uncertainty in untransformed sulfide concentration generated by uncertainty in voltage measurements varies with the magnitude of the uncertainty in the millivolt readings whereas the relative uncertainty does not vary with concentration. For millivolt errors of ±0.1 mV the corresponding sulfide values have an absolute difference of 0.8 to 78 μM from the true value and a relative difference of 0.8%. For millivolt errors of ± 1 mV the corresponding uncertainties are an order of magnitude larger than ± 0.1 mV. The upper and lower sulfide values now have an absolute difference of 7.8 to 781 μM from the true value and a relative difference of 7.8%. This is the value indicated by Rundle (2011) as being typical for ISE measurements of divalent ions such as sulfide. Unlike the differences in transformed sulfide the absolute difference increases with concentration and the relative difference remains constant for a given millivolt uncertainty.

Unlike the situation for transformed sulfide there is a positive bias in untransformed sulfide. This occurs because the nonlinear, i.e., logarithmic, nature of the calibration relationship generates an asymmetry in the back-transformed sulfide estimates relative to the true value. When the absolute bias is defined as the average of the upper and lower concentrations associated with the millivolt uncertainty minus the true concentration the bias is seen to increase with the millivolt uncertainty for a given sulfide concentration and with the sulfide concentration for a constant millivolt uncertainty (Appendix 6). The relative bias, defined as the absolute bias divided by the true concentration, also increases with the millivolt uncertainty and unlike the absolute uncertainty, the relative bias is independent of concentration. In the ±0.1 and ±1 mv cases the relative bias is small, i.e., less than 1%. However, as the error in millivolts increases the relative bias increases so that a millivolt error of 10 mV results in a positive bias of 32% in sulfide. A millivolt error of this magnitude is unlikely.

Having explored the potential influence of millivolt error on sulfide error, it is useful to consider what expected errors in millivolt readings might be.

The voltmeters used in association with the ISE sulfide method often show the voltage potential using digital displays that show values to one decimal point. A first estimate of the random component of measurement uncertainty can therefore be assumed to be ± 0.1 mV (Taylor, 1997), although this may be an artifact of the digital display rather than a true estimate of meter accuracy. This estimate of uncertainty does not include bias of the voltmeter, it only includes non-systematic uncertainty; any voltmeter bias that may exist is compensated for in the calibration parameters. Assuming the display reflects a rounded value that is normally distributed, the ± 0.1 mV may be considered to indicate that 95.4% of repeated measurements would occur within $\mathrm{mV} \pm 0.1$. The corresponding estimate of relative random uncertainty in each sulfide concentration derived from the calibration relationship, when it is assumed there is no error associated with the nominal concentrations, is calculated to be 0.99 to 1.01 (99% to 101%).

In practice the millivolt readings associated with a sediment sulfide sample vary through time with the reading depending on when the reading is taken. This temporal variation is the consequence of the response time of an electrode not being instantaneous and the sulfide concentration in the sample not being in equilibrium. The response time of electrodes varies between electrodes and electrode types. In general ISE electrode systems require three to four minutes to reach a completely stable reading and less than thirty seconds to come within one to two millivolts of the final stable reading (NICO, 2000). The response time of the ThermoScientifc electrode is such that 99% of the stable reading should be reached within one minute or less for concentrations in the linear range of the electrode (TFS 2007). These times apply when the ionic concentration in the sample is not changing. Readings that indicate concentrations are changing after the response time, i.e., about 1 minute, should be considered carefully, since the readings may no longer be indicating the concentration of the sulfide in the undisturbed pore water. The changing readings may be indicating that sulfide complexes within the sediment are breaking down and transferring sulfide to the pore water. Many authors have reported that stable readings are achieved within 2 minutes but not always within 1 minute (Lewis-McRae pers. comm.).

In addition to millivolt stabilization, the variation around stable millivolt readings varies. Measurements made on standard solutions of various sulfide concentrations by one laboratory that used a single ISE probe and meter ranged from 0.3 to 1 mV (Chang, et al., 2014). In contrast, standard deviations calculated from measurements made by another laboratory on the same standard solution concentrations but using several different ISE electrode and meter combinations obtained a range of 1.1 and 1.7 mV (Chang, et al., 2014).

In practice the variation in mV readings may therefore be more on the order of 1 mV. The effect of the 1 mV error in millivolt readings on the frequency distribution of $\log_{10}(S)$ and S values is illustrated in Figure 2. Details of how the numbers were calculated are given in Appendix 6. Note the distribution of the ISE error or uncertainty around the calibration curve is Gaussian and hence, the distribution of $\log_{10}$S is normal whereas the distribution of the untransformed sulfide is skewed by the non-linear nature of the inverse relationship and is log normal.

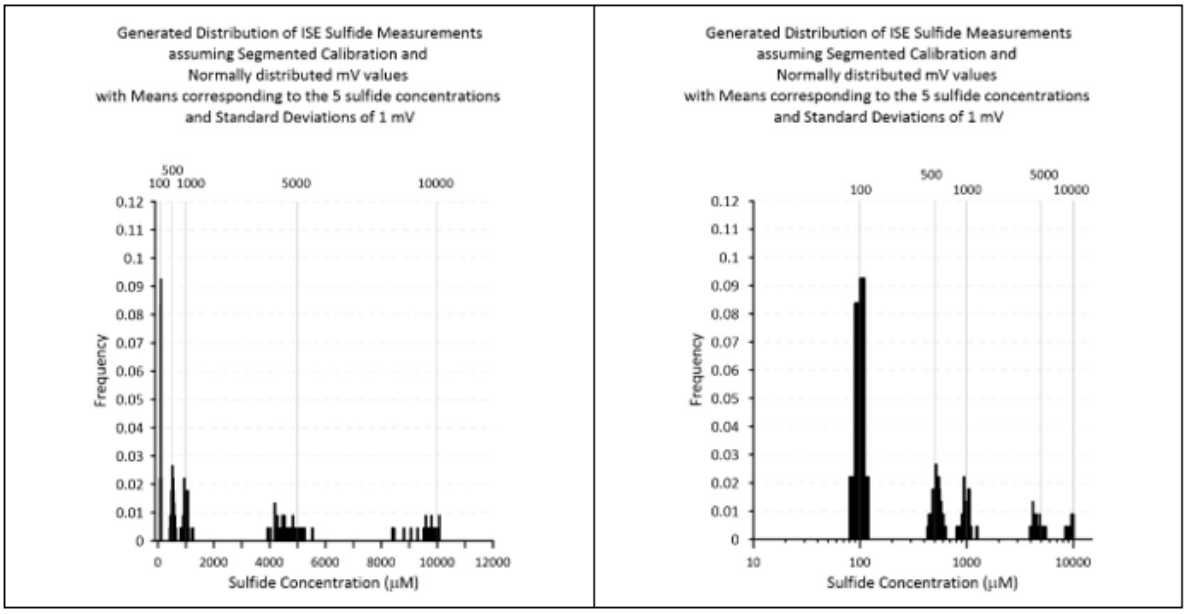

Figure 2: Illustration of the effect of a 1 mV error on sulfide (left panel) and log_{10} transformed sulfide (right panel) values. The error is assumed to be normally distributed with a mean of zero and a standard deviation of 1 mV.

As indicated earlier, ISE based measurements are most useful when order of magnitude measurements are needed. However, the method can achieve accuracy and precision levels of ± 2 or 3% when used very carefully, probes are properly and frequently calibrated, and analysts are well aware of the method limitations (Rundle, 2011). If variation in voltage readings is of order 1 mV, the theoretical error in the measurement of the concentration of a bi-valent ion such as the sulfide ion (S^{2-}) when measurements are made at 25°C is expected to be about 8% (Rundle, 2011 and Appendix 6).

Uncertainty Associated with Standard Solution Errors

The error associated with the titration-based estimates of standard concentrations can be considered to be about 0.1% when careful technique is used (TFS, 2007).

The Aquaculture Monitoring Standard 2018 associated with the Canadian Aquaculture Activity Regulations requires the use of an ISE electrode and meter that achieves a 5% accuracy or better (DFO, 2018). The above considerations suggest that errors in millivolt readings may be sufficient to exceed this limit and that in order to achieve the required regulatory accuracy precise calibration combined with careful and consistent analytical technique must be followed.

In practice, the total error depends on multiple factors in addition to the millivolt error. Some of these factors are:

- the slope of the calibration relationship,
- temperature of the sample verses temperature of the calibration solutions during calibration,
- deviations of the sample and probe from "ideal" behaviour,
- impurities or contamination of the ion-selective membrane,
- whether readings are near the detection limit of the electrode,
- whether readings are within the linear range of the calibration relationship, and

- if the slope remains constant over the range of concentrations and the time period required for the analyses.

EMPIRICAL EVIDENCE

The above paragraphs examined the potential for uncertainty in sulfide estimates based on a consideration of the propagation of input errors. In the following paragraphs we examine some empirical evidence that estimates the uncertainty associated with sulfide measurements. We use the terms uncertainty, bias and precision and assume uncertainty or accuracy includes bias and precision errors, precision is the random or non-systematic scatter of measurements around the true but unknown value and bias is the systematic error causing the center of the scattered measurements to deviate from the true value.

Empirical insight into the uncertainty of the ISE method can be obtained by examining the variability in measurements made on solutions with known sulfide concentration. Chang et al. (2014) prepared serial dilutions of a sulfide stock solution to generate a series of samples with "known" sulfide concentrations. A subset of these samples was used to generate ISE calibration relationships between millivolt readings generated by sulfide ISE measurements and "known" or nominal sulfide concentrations. The measurements were made using the Accumet P125 with a variety of sulfide electrodes. In all of the Chang et al. (2014) experimental results considered in this section, stock sulfide solutions were freshly made, dilution samples were not stored, calibration relationships were derived just prior to sample measurements and measurements were taken immediately after the addition of SAOB. The experiments were conducted by personnel based in two different laboratories and involved multiple ISE probes and voltmeters. The measurements provide a set of measured and "known" sulfide concentrations that can be used to empirically estimate the accuracy of the ISE method recognizing that the derived accuracies are from the least variable of situations, i.e., no potential confounding sediment pore water interactions such as complexing with sediment matrix components, sample storage etc.

Precision

Estimates of measurement precision, both absolute and relative, derived from the data reported in Table F6a, F6b and F8 of Chang et al. (2014) are shown in Figure 3. Absolute precision was estimated by calculating the sample standard deviation (SD, absolute precision) and the coefficient of variation (CV, relative precision) of the measurements. The sample size associated with each estimate is ten. These values indicate the non-systematic error or variability in the measurements relative to the mean values of the experimental setups. The standard deviations varied from 2 μM to 1176 μM with the SDs increasing with the mean and nominal concentration. The CVs varied from 1.7 to 11.6% and were highest for the largest concentrations. The CVs were less than 10% for the 100, 500, 1,000, and 5,000 solutions and were reasonably independent of the mean and nominal concentration. The mean and median CVs for all concentrations were respectively 4.6% and 3.9%. The frequency distribution of the CVs is also shown in Figure 3. Seventy percent of the CVs are less than 5% and thirty percent are greater than the 5% prescribed by the Canadian Aquaculture Monitoring Standard (DFO, 2018). The magnitude of most of the relative precision estimates are consistent with the more theoretical error propagation considerations. It is unlikely that the level of precision can be improved since they probably represent an underestimate of the precision associated with in situ sediment samples since they do involve sediment collection, processing, and sediment matrix effects. These estimates of precision do not account for bias, i.e., deviations from the true or nominal value.

Bias

Estimates of absolute and relative bias, sometimes referred to as trueness, derived from the above data are shown in Figure 4. The trueness or absolute bias is the difference between a mean ISE value and the expected or ‘true’ value, which in this case is assumed to be the nominal value, and the relative bias is the absolute bias divided by the ‘true’ value. The relative bias varied from -10.9% to 10.4% with the mean and median biases being respectively -0.9% and -1.2%. In the frequency distribution of the relative biases, seventy percent of the biases were negative i.e., underestimated the true mean and 30% were positive and overestimated the true mean. Twenty percent of the relative biases were within ± 1% of the true value and 18% were greater than 5%. The bias is therefore not consistent and cannot be easily accounted for since it varies with electrode, concentration, and analyst, among other things. However, with a concerted effort the degree of bias for a laboratory and analyst might be established if after many tests, the bias is consistent.

Uncertainty and Accuracy

The most relevant indicator of uncertainty includes the effects of bias (trueness) and precision. This combined indication is shown in Figure 5 by the root mean square error (RMSE) and the relative RMSE. The RMSE is essentially the SD calculation with the deviation calculated as the difference between the measured values and the true value rather than the sample mean. The relative RMSE is the RMSE divided by the true value converted to a percentage. The relative root mean square error varied from 1.8% to 13.7% with the mean and median biases being respectively 5.9% and 5.0%.

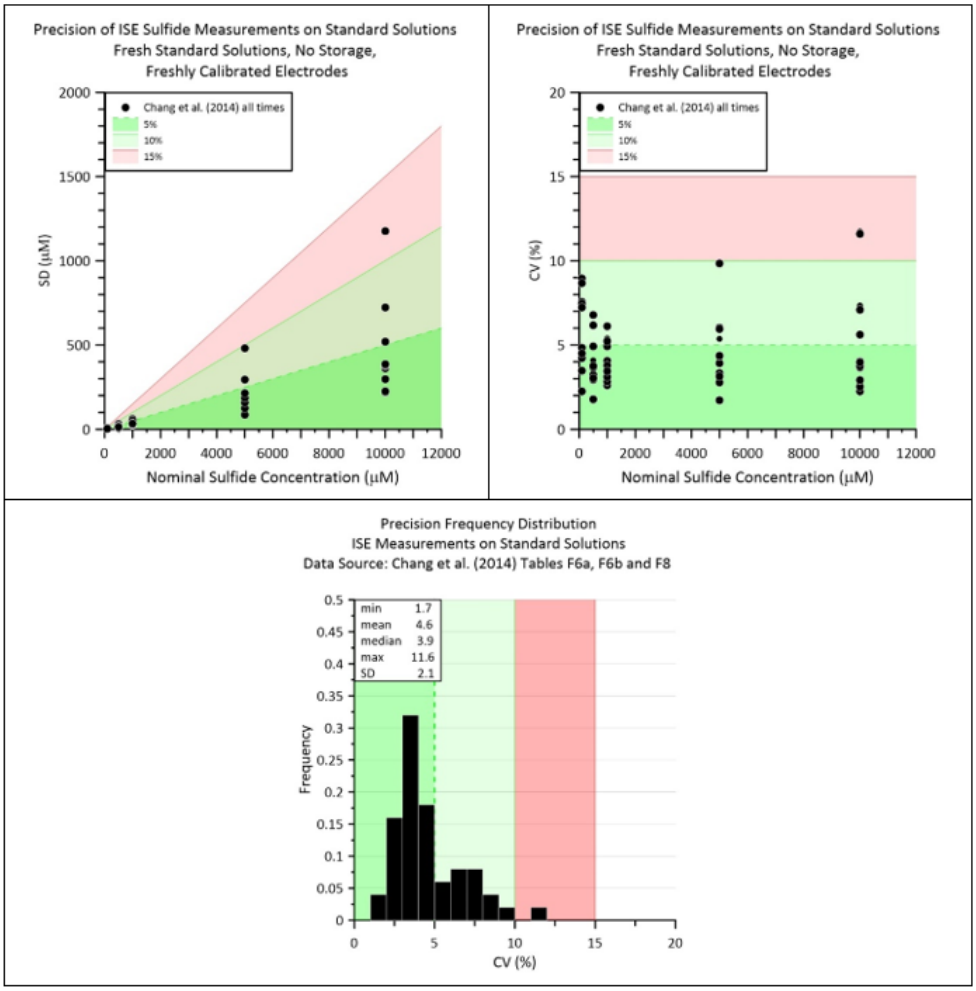

Figure 3: Estimated precision in ISE sulfide measurements derived from data reported in Chang et al. (2014). Measurements were taken on prepared aqueous solutions. The nominal sulfide concentrations of the solutions were determined from titrations. The solutions were not stored. All precision estimates were calculated from sample sizes of n=10.

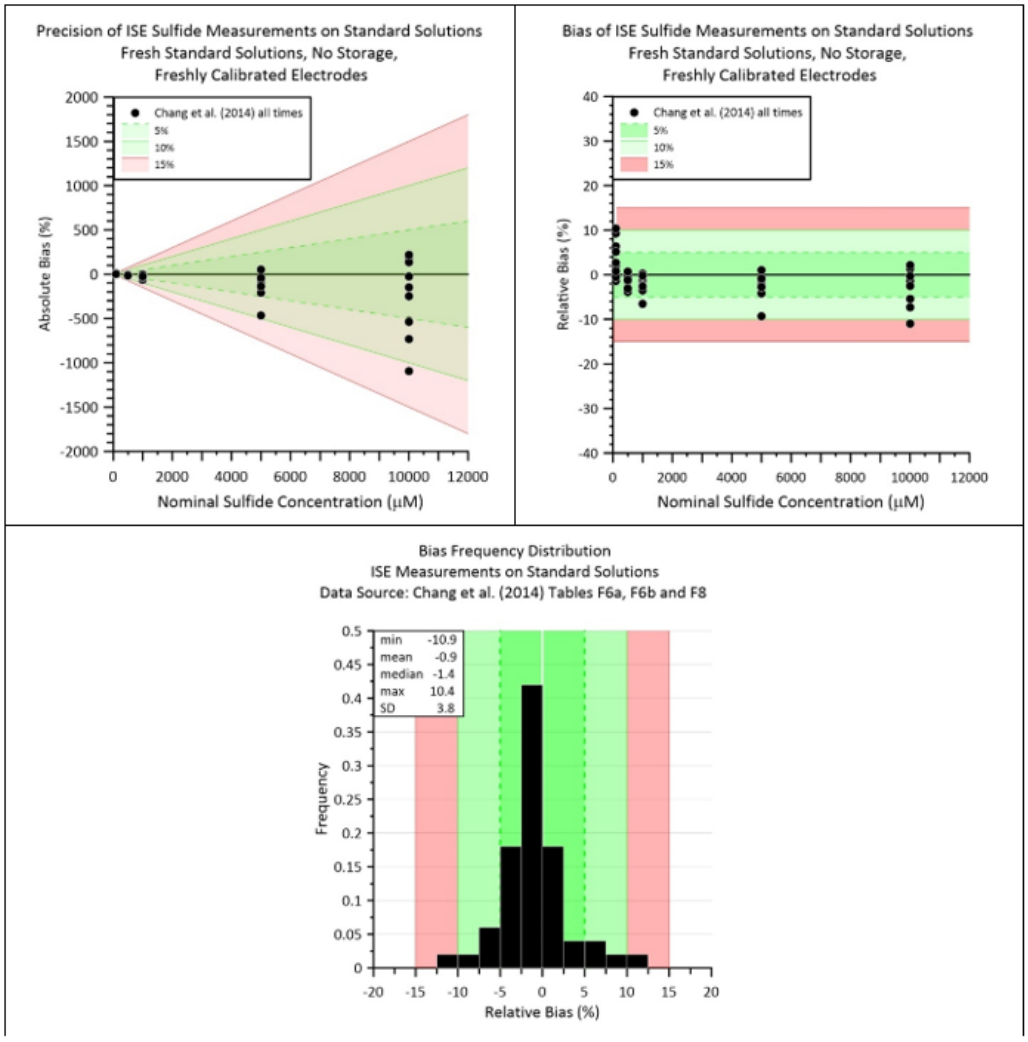

Figure 4: Estimated bias in ISE sulfide measurements derived from data reported in Chang et al. (2014). Measurements were taken on prepared aqueous solutions. The nominal sulfide concentrations of the solutions were determined from titrations. The solutions were not stored. All biases were calculated from sample sizes of n=10.

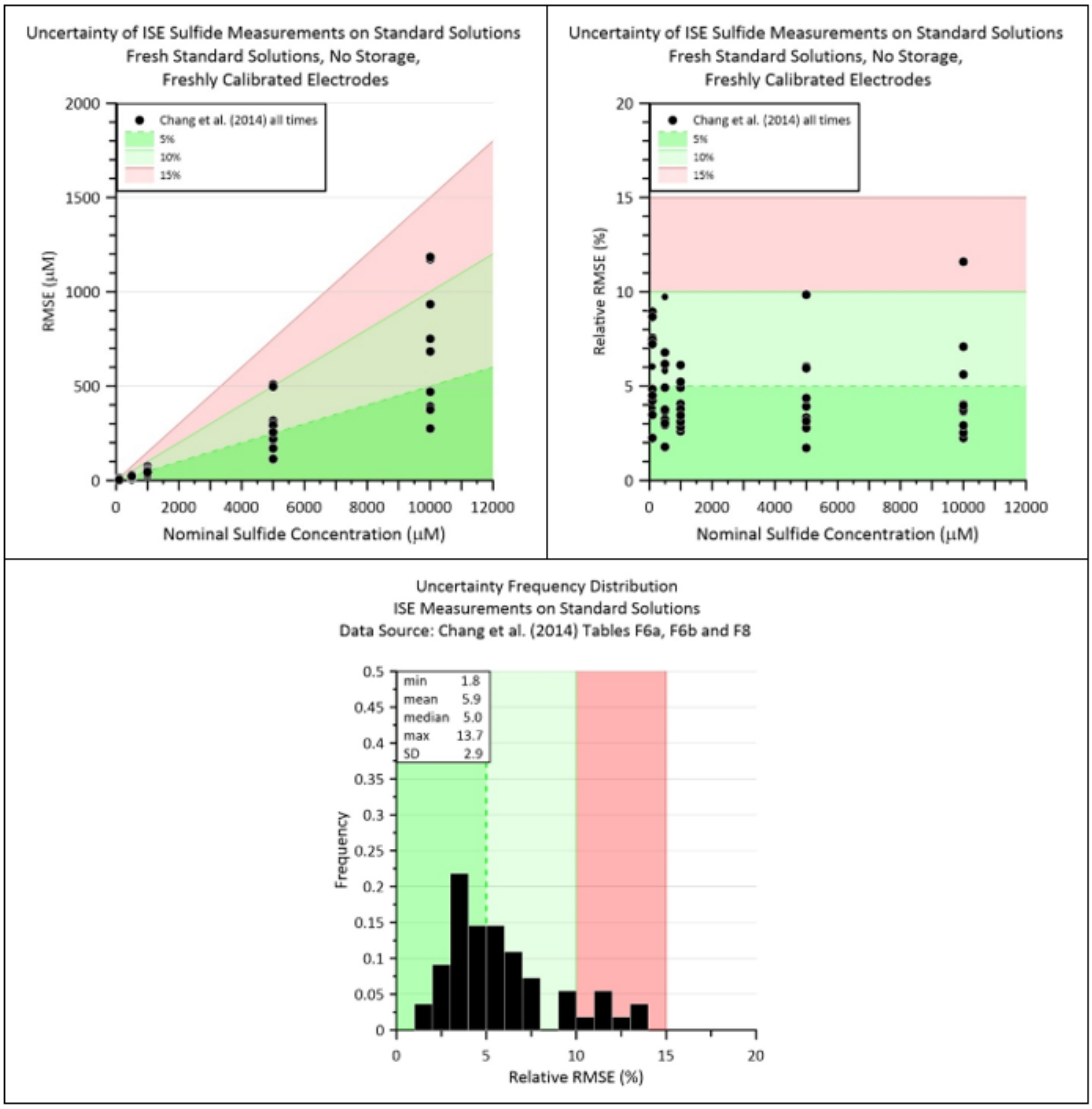

Figure 5: Estimated total uncertainty in ISE sulfide measurements derived from data reported in Chang et al. (2014). Measurements were taken on prepared aqueous solutions. The nominal sulfide concentrations of the solutions were determined from titrations. The solutions were not stored. All uncertainties were calculated as RMSE with deviations calculated as differences between measured and true values. In all cases the sample size was n=10.

The above considerations of ISE uncertainty provide a baseline indication of the accuracy of the ISE method under ideal conditions. Deviations from these conditions increases the uncertainty. Several of the factors resulting in deviations from the baseline are considered below.

SECTION 3: FACTORS AFFECTING ISE UNCERTAINTIES

Many factors have the potential to influence ISE measurements. These include factors associated with the field collection of samples, the storage of samples between collection and analyses, the laboratory analyses of the samples as well as in situ conditions affecting the in situ sediment matric. Some of the factors are:

- exposure time in air after sediment collected and subsampled and mixed with L-ascorbic acid and SAOB (Hargrave et al., 2008b)
- time elapsed between L-ascorbic acid and SAOB addition and measurement reading (Hargrave et al., 2008b), although this effect can be minimized by taking readings within 2 min of addition
- water content of sediment sample (Hargrave et al., 2008b)
- the nature of the sediment matrix including sediment type, feed and feces content, bacteria
- temperature differential between the sediment sample being analysed and the temperature of standard solutions used to calibrate the ISE electrode and meter system (Hargrave et al., 2008b)
- accuracy of the standard solutions
- accuracy of the millivolt reading
- skill and experience of the analyst (Schaaning & Hansen, 2005) and demonstration of the good laboratory capabilities (Lewis-McCrea pers. comm)

Some of these factors have been considered earlier in this document and in the accompanying documents (Wong and Page 2025a,b). We therefore only consider uncertainties associated with some of the influencing factors here. We attempt to quantify some of the uncertainties associated with the ISE methodology and show how the uncertainty increases from the magnitudes typically achieved in close to ideal and well controlled situations to those experienced in operational field measurements that incorporate the magnitudes of spatial and temporal variations in sulfide measurements.

TIME FOR ISE READING TO STABILIZE

Once L-ascorbic acid and SAOB have been added to a sample, the combined slurry is temperature controlled and stirred and the ISE electrode has been inserted, the millivolt reading will evolve. In many cases the time between stable millivolt readings and insertion of the ISE electrode into a sample plus SAOB occurs within 1 to 2 minutes (Hargrave et al., 2008a; 2008b). It should be noted, however, that if the electrode and stirring continues and/or a millivolt reading is made at some time after this, particularly after 5 min, the millivolt reading may differ from the initial ‘stable’ reading and the value of the measurement should be considered carefully (Hargrave et al., 2008b). This is because the chemical equilibrium within the slurry has not been established, i.e., processes of solubilization of solid phase sulfides were not complete, and the millivolt meter often defines stability as less than a specified magnitude of change within a specified number of seconds.

In the original development of the method, it was thought that the effect of solubilization could be minimized by making measurements immediately after addition of SAOB (Hargrave pers. comm.). However, as the evidence reviewed here indicates, it is now acknowledged that this assumption was not correct (Hargrave pers comm.). Never-the-less, taking readings within the initial 1 to 2 minutes after addition of SAOB helps reduce the effect of solubilization and is still preferred approach for measuring the in situ pore water ‘free’ sulfide using the ISE method and this operational criterion has been incorporated into ISE protocols (Wong and Page, 2025a,b). Clearly readings taken at time periods longer than this should be avoided (Hargrave et al., 2008b).

AGE OF ISE ELECTRODE

The slope of ISE calibration relationships typically changes over time for multiple reasons including aging of the probe and probe contamination through use in destabilizing environments (Rundle 2011). This change can result in a reduction in measurement accuracy (Rundle 2011). Laboratories can make efforts to minimize the error associated with these changes by keeping track of probe calibration performance, drift rates etc. (Lewis – McCrea pers. comm).

STORAGE OF ELECTRODES OR ELECTRODE DRIFT

One of the factors influencing the accuracy of ISE measurements is the time elapsed since electrode calibration. In principle one would expect the uncertainty of measurements made with stored electrodes to increase but the magnitude and time needed for the effect to be noticeable are not so obvious. According to Rundle (2011) the slope of ISE calibration relationships typically changes, and more specifically reduces, as the electrode ages and gets contaminated and this change results in a reduction in measurement accuracy (Rundle 2011).

Fortunately, the magnitude of the effect has been examined. Chang et al. (2014) made new sulfide standard solutions at time intervals of 0, 1, 2, 3, 4, 12, 24, 48, 72, and 96 hours after the time zero solutions were made. ISE sulfide determinations were made using electrodes that were calibrated just prior to making measurements and electrodes that were calibrated at time 0 h and reused at subsequent times without recalibration. The results for the freshly calibrated measurements were presented in Section 2 in relation to base uncertainty. The results of the measurements made with the unrecalibrated or stored electrodes were included in Appendix F of Chang et al. (2014) and a reanalysis in the context of accuracy and precision are presented here.

The precision, as indicated by the CV for electrodes used soon after (0h) calibration was approximately 5% or less and was consistent with the empirical evidence for the time zero control CVs present in the previous section (Figure 6). The CVs for storage times of 1 to 4h were variable and centered around 5% whereas the CVs associated with storage times of 12 to 96h were consistently greater than 5% with some exceeding 10% and one exceeding 15% (Figure 6). With the exception of 100 μM and 500 μM nominal concentrations, the CVs tended to increase with storage time Figure 5. For nominal concentrations $\leq$1,000 μM, the CVs tended to decrease with nominal concentrations whereas for nominal concentrations above $\geq$1,000 μM the CVs tended to increase with nominal concentration (Figure 6).

The effect of electrode storage is to increase bias relative to the baseline bias Figure 7. Baseline biases were for the most part less than 5% (also see Section 2). Only the electrode biases associated with 0, 1, and 2 h electrode storage times were consistently less than 5%. Storage times between 2 h and 4 h resulted in bias between 5 and 10% and storage times greater than 4 h resulted in bias greater than 10% and for the long storage times greater than 20%. These may be underestimates of the magnitude of the effect since the probes used to examine the bias were relatively new.

Total uncertainty relative to the baseline uncertainty (Figure 8). The total uncertainty for measurements made with electrodes that had not been stored, i.e., a storage time of 0 h, varied from 2.7% to 6.5% with a median of 4.7%. The uncertainty for measurements with electrode storage time of one to two hours increased to between 3.6% and 7.4% with a median of 5.6 h and the uncertainty varied from 5.9% to 11.3% for post calibration times of 3 to 4 h. The uncertainty increased to above 10% for post calibration times of 12 h or more.

The above analyses indicate ISE electrodes lose their calibration as post-calibration time increases. The sulfide measurements decrease in precision, as evidenced by an increase in

SDs and CVs, become more negatively biased i.e., measurements underestimate the true sulfide value, resulting in an increase in total sulfide measurement uncertainty (reduction in accuracy). At post-calibration storage times greater than 2h the median bias increases to greater than 5% and after 4h exceeds 10%. The total uncertainty is increased from a baseline at zero hours post calibration of generally less than 5% to more than 5% at post calibration times of 3h or greater. The increases in uncertainty can be avoided by recalibrating electrodes every 2-3 hours.

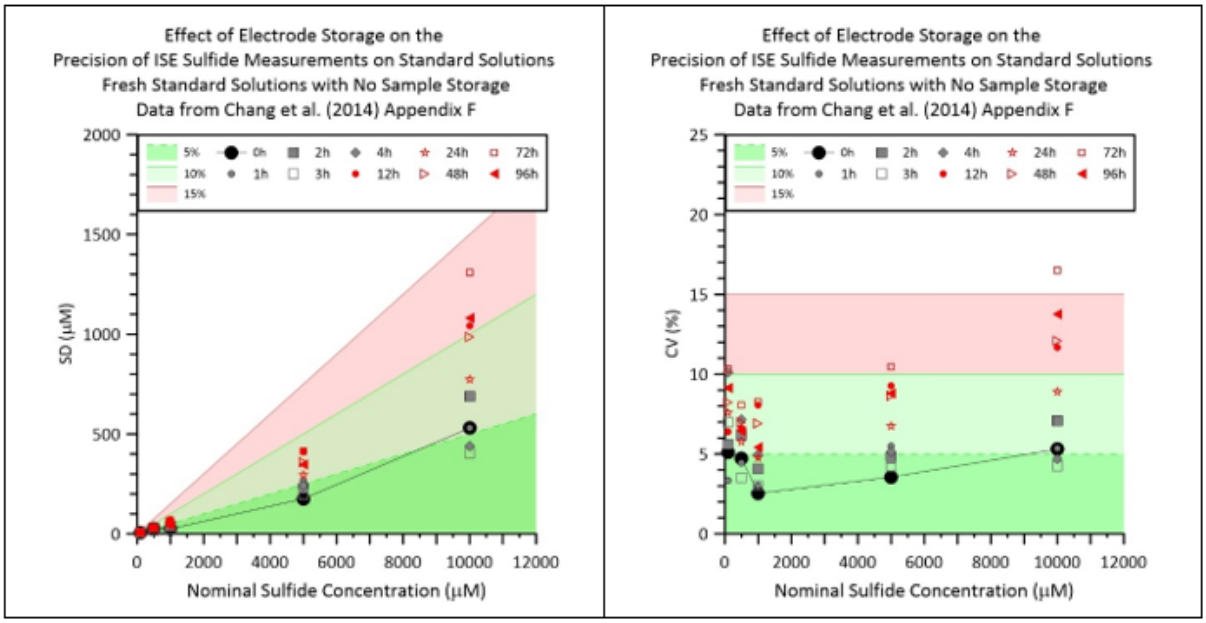

Figure 6: Estimated precision in ISE sulfide measurements made on standard sulfide solutions in relation to electrode storage time. The precision estimates are derived from data reported in Chang et al. (2014) Appendix F. The nominal sulfide concentrations of the solutions were determined from titrations. The solutions were not stored. All calculations were based on sample sizes of n=12. The black, grey, and red symbols indicate storage times of 0h, 1-4h and 12-96h respectively.

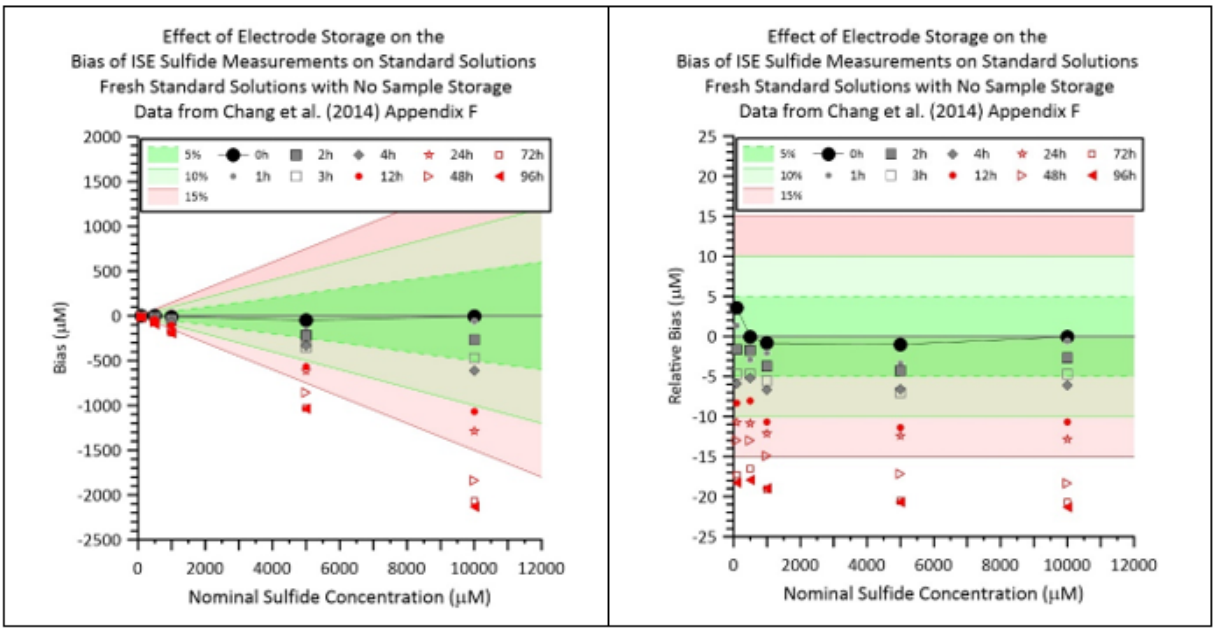

Figure 7: Estimated bias in ISE sulfide measurements made on standard sulfide solutions in relation to the time since electrodes were calibrated i.e., the electrode storage time. The bias estimates are derived from data reported in Chang et al. (2014) Appendix F. The nominal sulfide concentrations of the solutions were determined from titrations. The standard solutions were not stored and the concentration of stock solutions was verified with lead perchlorate titration. All calculations were based on sample sizes of n=12. The black, grey, and red symbols indicate storage times of 0 h, 1-4 h, and 12-96 h respectively.

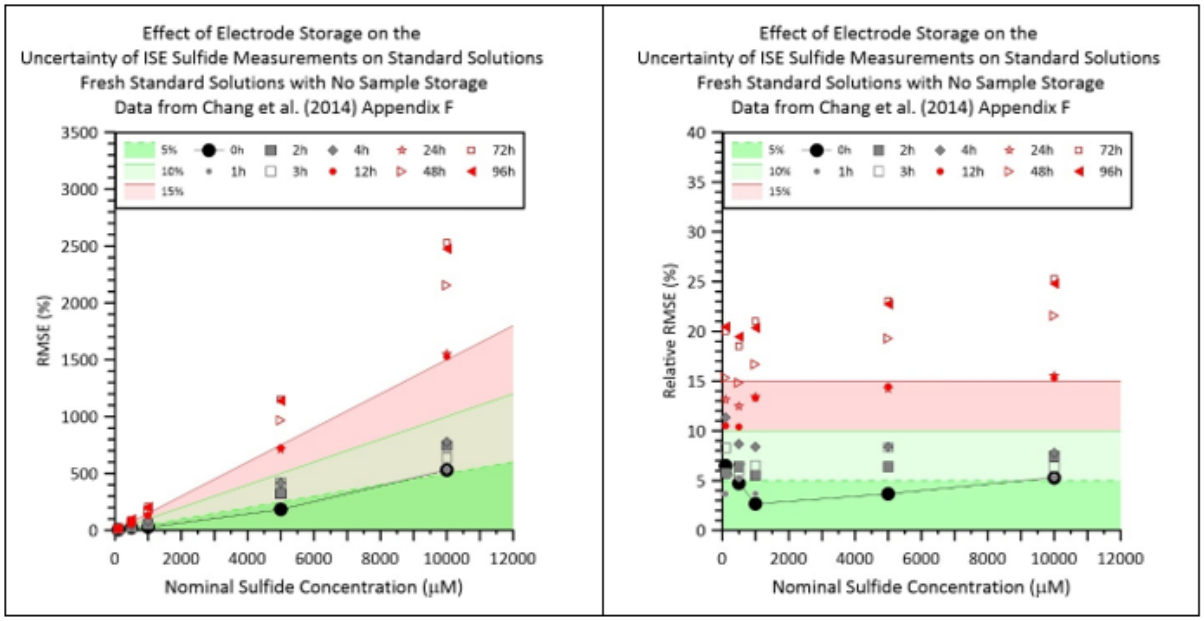

Figure 8: Estimated total uncertainty in ISE sulfide measurements made on standard sulfide solutions in relation to nominal sulfide concentrations and time since electrode calibration i.e., the storage time. The uncertainty estimates are derived from data reported in Chang et al. (2014) Appendix F. The nominal sulfide concentrations of the solutions were determined from titrations. The nominal solutions were not stored. All calculations were based on sample sizes of n=12. The black, grey, and red symbols indicate respective storage times of 0h, 1-4h and 12-96h.

STORAGE OF LABORATORY STANDARDS AND FIELD SAMPLES

The effect of sample storage on ISE estimates of sample sulfide concentration has been considered in detail in (Wong and Page, 2025b).The perspective of how the bias and precision of ISE measurements varies with time stored is highlighted here.

Aqueous Standards

Figure 9 shows the how the relative bias and coefficient of variation varied with storage time of aqueous standard sulfide solutions. The data shown are from the laboratory experiments described by Chang et al. (2014). The experiments were conducted by staff at two separate laboratories. All experiments were conducted on standard solutions of sulfide, i.e., none of the experiments included sediment in the samples. All of the solutions could be considered as homogenized and the measurement error (uncertainties) associated with the measurements does not include sediment matrix effects.

The bias was predominately negative with only two samples showing positive bias (one dot on top of the other) and these were both for very high and largely unrealistic values of sulfide (20,000 and 30,000 µM) Figure 9.

The coefficient of variation was variable and showed a trend of increasing with storage time. For storage times up to 24 h the magnitudes of the CVs were centered around 5% and were generally less than 10%; CVs increased considerably after 24 h of storage.

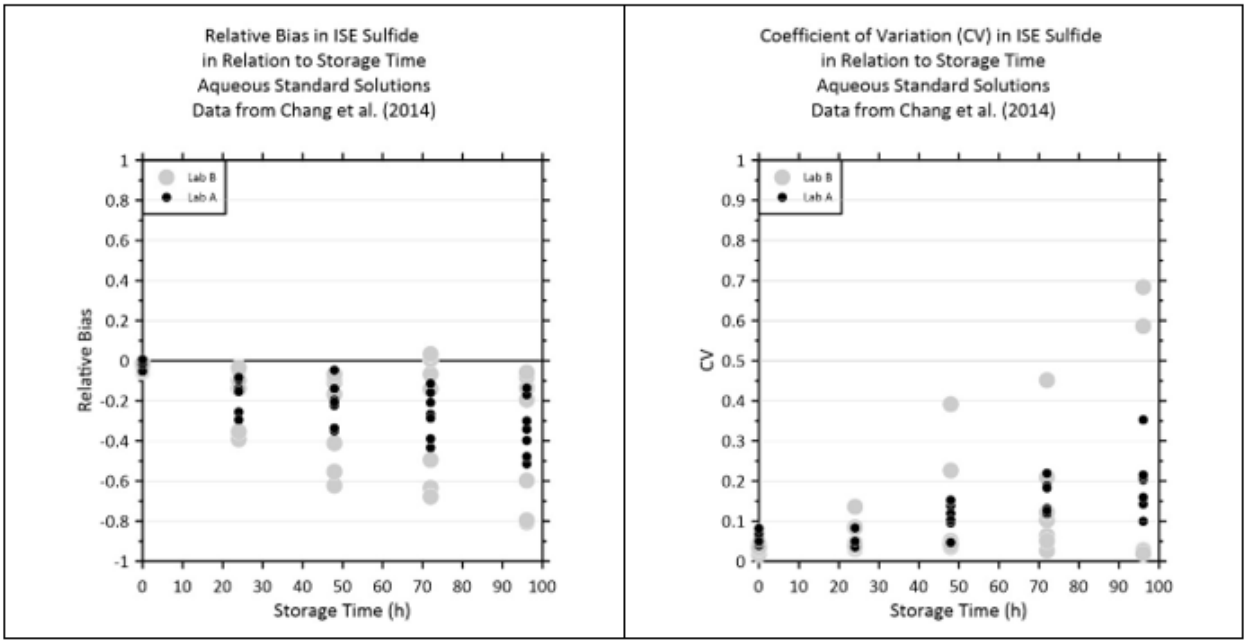

Figure 9: The variation in relative bias and precision (CV) of ISE measurements made on stored aqueous standard solutions. The data is associated with the storage stability experiments described in Appendix E of Chang et al. (2014).

Sediment

Figure 10 shows how the relative bias and coefficient of variation varied with storage time of sediment samples collected from several locations. In the case of Wildish et al. (1999) samples were collected from two salmon farm sites in the Quoddy Region of southern New Brunswick and one southern New Brunswick intertidal site. Some of the samples collected were

homogenized prior to analysis and some were left unmixed. In the case of Wong and Page (2025b), samples were collected from locations near two open net pen Atlantic salmon (*Salmo salar*) farm cages in the Lime Kiln Bay area of the southern New Brunswick area of the Bay of Fundy. The sediments were collected with a small Hunter-Simpson benthic grab. Multiple grabs from each location were collected, combined in a bucket and homogenized using a powered mechanical paint mixing device which would have introduced some oxygen into the sediment. Three subsamples of homogenized sediment were taken at nine time intervals (0, 3, 6 , 24, 48, 72, 96, 120, and 144 h) after collection using 5-ml plastic syringes. The subsamples were analyzed for total free sulfide using the ISE method described in Wong and Page (2025a). The biases were calculated by assuming the true value was the first measurement. The reference storage times of 36 h and 72 h are the limits suggested respectively by the DFO Aquaculture Regulations (DFO, 2018) and by Wildish et al. (1999). The 4 h storage time is provided as a shorter time reference.

There was some degree of bias in almost all ISE measurements taken post initial measurements (Figure 10). Both Wildish et al. (1999) and Wong and Page (2025b) observed positive and negative bias, although the bias was predominately negative; five of the eight experiments showed negative bias and three of the eight experiments showed positive bias. Wildish et al. (1999) suggested there were differences between the initial ISE sulfide measurements, the C_{To} concentrations, and the measurements made at 3 h (C_{T3}) although the evidence was not strong, in part due to limited sample sizes. They found the differences were statistically significant ($p<0.05$) for samples stored in vials but not for those stored in syringes. However, they also concluded that the difference between ISE sulfide measurements made at identical storage times resulted in no statistically detectable differences between the sediments stored in vials and syringes. Wong and Page (2025b) also found that sulfide measurements varied with time (Figure 10, red symbols). They found that measurements made after the first measurement (t0) differed from those at t0 by 5% to ~40%, with the differences varying between samples and sediments. The differences were not statistically significant until a storage time of 48 h (Wong and Page 2025b). Although the data indicates considerable variation in the relative bias of the ISE sulfide measurements it should be remembered that the bias is relative to the first measurement in a storage time series and not relative to the time the sample was taken. As such the biases are likely underestimates of bias relative to a measurement taken as soon as the sediment is sampled since the sulfide concentrations seem to begin changing as soon as they leave the in-situ situation. It is also of interest that the positive biases were associated with highly organic enriched sediments. Although all of the experiments were conducted using small sample sizes ($n<5$) these sample sizes are similar to those required for Tier 1 regulatory sampling in New Brunswick (Chang et al., 2013b).

Of note, the positive and negative spread of the bias in the measurements made on stored sediments differs from the almost exclusively negative bias associated with stored standard solutions, although the range in magnitude of bias is similar in both cases, about 0.0 to 0.8.

Clearly, the composite plot of bias (Figure 10) indicates there is little consistency to the bias associated with stored sediment, so a simple correction cannot be made. The data suggests the frequency distribution of the bias is approximately normally distributed with a mean overall relative bias of – 0.11 (-11%) and a standard deviation in the bias of 0.36 (36%). Without many more experiments and an ability to group sediments by easily identified sediment characteristics it appears it might be prudent to consider the bias as part of a random process with central tendency and a standard deviation. This is unlikely to be satisfactory for robustly detecting impact signals of order 1,000 μM, especially when small sample sizes are involved.

The coefficient of variation varied between almost 0.0 to 0.5, although most values were less than 0.2 (Figure 10). The central tendency for CVs was about 0.1 (10%) and there appeared to

be no apparent trend with storage time. These are higher than the CVs obtained for measurements made on aqueous standard solutions in the first 24 h of storage. The lack of trend with storage time differed from the results associated with aqueous solutions which showed an increase in CV with storage time. These differences are presumably associated with the more complex nature of the sediment matrix relative to the matrix of the aqueous stand solutions.

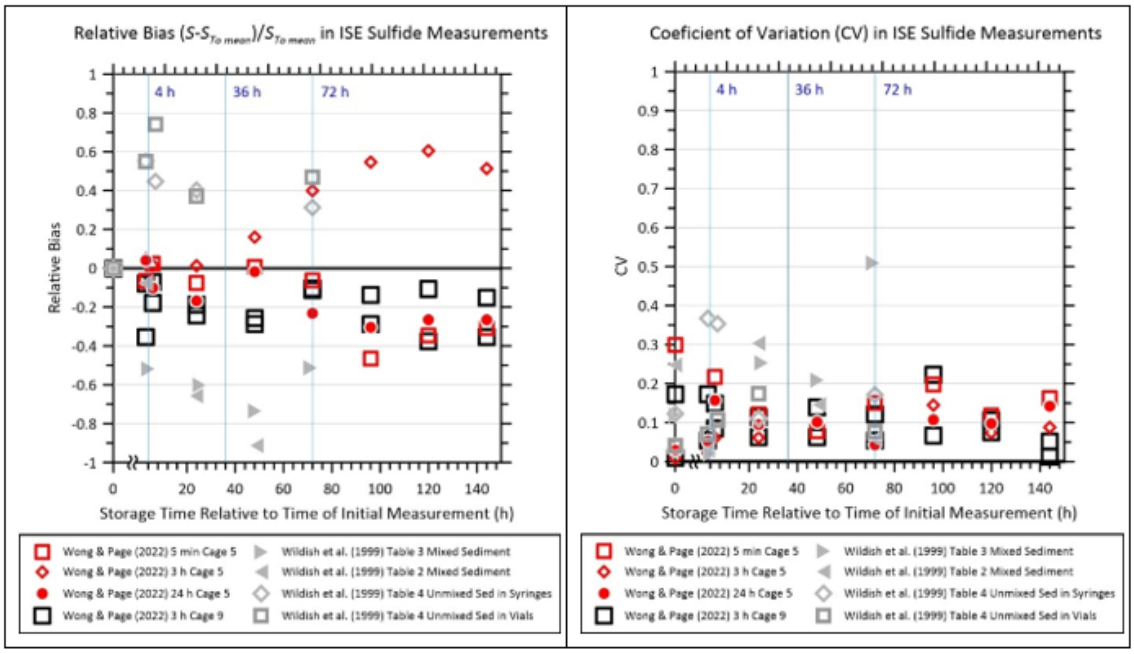

Figure 10: The variation in relative bias and precision (CV) of ISE measurements made on stored sediment samples. The data are associated with the storage stability experiments described in Wong and Page (2025b) and Wildish et al. (1999). The vertical lines represent reference points for storage durations. All of the Wong and Page (2025b) results are associated with mixed sediment. The times mentioned in the legend are the time of t0 relative to sediment collection. The time indicated on the x axes is the time relative to the first sulfide measurement.

SEDIMENT WATER CONTENT

The ISE method is based on establishing a calibration relationship between a millivolt reading generated by an ion selective electrode immersed in an aqueous solution of known sulfide concentration and connected to a voltage meter. The known concentrations are made by diluting a stock solution of sulfide by known amounts and adding a volume of pH buffer solution, the L-ascorbic acid and SAOB, to each of the concentration standards. This addition effectively dilutes the standard concentration and the ISE millivolt reading actual reflects the ion activity of the diluted solution. The ion activity and millivolt reading changes with the dilution ratio. The dilution effect is accounted for in the calibration relationship because the relationship is between the millivolt readings and the assumed standard concentration, not the actual concentration. Using the calibration relationship to estimate or measure the sulfide concentration within a sample is therefore most accurate when the ratio of the volume of sample solution containing the sulfide to the volume of the buffer added is the same as the ratio of sample volume to buffer volume used in the establishment of the calibration relationship. The ratio of standard or sample volume to buffer volume recommended by Wildish et al. (1999) and Wildish et al. (2004a) is 1:1 and this is the ratio that has been used by many, including Chang, et al. (2014). Regulatory documents often repeat the necessity of maintaining a 1:1 ratio between the volume of sediment

being analyzed and the volume of L-ascorbic acid and SAOB; however, the same documents do not mention corrections for water content (NBDELG, 2004; NBDELG, 2018).

The dilution ratio of 1:1 can easily be maintained, when the sample is water, by ensuring the volumes of sample water and buffer are the same. However, when the sample is sediment, the ratio is not easily maintained because the volume of the sediment sample is not all water. The following paragraphs discuss the influence of variations in the dilution ratio to estimates of the pore water sulfide concentration.

The amount of water within sediments, i.e., the proportion of the sample volume that is water (the pore water), varies between sediments, with the proportion being highly correlated with sediment grain size (Hargrave et al., 2008a). The water content is typically higher in fine grained than in coarse grained sediments (Hargrave et al., 1993a).

The water content of coastal marine sediments in geographic areas where marine aquaculture sites are located, is quite variable. The water content of individual samples varies from 12% to 79.8% (Table 2). The geographic average or mean water content varies between about 32.8% to 64.8% (Table 2). The standard deviation of values within an area varies from 1.0% to 17.4% (Table 2). According to Nova Scotia open data the range may be even larger, being 13.4 to 91.1% (Lewis McCrea pers. comm.).

The water content in sediment samples taken from locations near open net pen fish cages and those from sediments collected at reference stations are similar in the sense that no consistent pattern has been detected; at least in the southwest New Brunswick salmon aquaculture area of the Bay of Fundy (Table 2, Hargrave et al., 1995; Hargrave et al., 1997). Sediment samples collected at stations adjacent to the edge of salmon net pens located in the Bliss Harbour, Fundy Isles region of the Bay of Fundy showed that water content varied (51% to 74%) between sampling dates and suggested there was little or no seasonal trend in the water content (Hargrave et al., 1993b).

Table 2: A limited summary of some sediment water content estimates for marine sediments in coastal areas used for finfish and/or shellfish aquaculture. Ranges are minimum reported values to the maximum reported values. SD stands for standard deviation.

Water Content (%)	Location and Context	Source
Range: 12.0 to 57.8 Mean: 35.9 to 47.1 SD: 3.5 to 4.8	Annapolis Basin region of the Nova Scotia Bay of Fundy	(Hargrave et al., 1993a)[3]
Range: 51 to 74	edge of the pens in Bliss Harbour, L'Etang Inlet, Bay of Fundy	(Hargrave et al., 1993b)
Range: 34.3 to 73.7 Mean: 57.2 SD: 12.3	Western Isles Region of the Bay of Fundy, New Brunswick Canada; **near open net pen** fish farms	(Hargrave, et al., 1995)[1]
Range: 18.9 to 71.8 Mean: 51.1	Western Isles Region of the Bay of Fundy, New Brunswick Canada; **reference** locations	(Hargrave, et al., 1995)[1]

Water Content (%)	Location and Context	Source
SD: 17.4		
Means: 32.8 to 64.8 SD: 1.0 to 16.3	Passamaquoddy Bay and Letang Inlet Southwestern Bay of Fundy	(Bugden et al., 2001)[2]
Range: 39.1 to 79.8	Tracadie Bay, PEI **within** suspended mussel culture **lease** areas	(Hargrave et al., 2008a)
Range: 33.4 to 75.9	Tracadie Bay, PEI outside of suspended mussel culture lease areas with no seagrass (*Zostera*) present	(Hargrave et al., 2008a)
Range: 21.9 to 50.7	Tracadie Bay, PEI outside of suspended mussel culture lease areas with seagrass (*Zostera*) present	(Hargrave et al., 2008a)
Range: 29.5 to 80.3	Winter Harbour, PEI	(Hargrave et al., 2008a)
Means (NB): 26.4 to 74.5 SE (NB): 1.2 to 7.7 Means (NS): 25.9 to 62.6 SE (NS): 1.2 to 23.7	in vicinity of a swNB finfish farm in vicinity of a NS finfish farm	data from B.D. Chang that was associated with (Chang et al., 2013)

[1] – The range, mean and standard deviation (SD) were calculated from the 0-2 cm sediment depth data provided in Appendix A of Hargrave et al (1995). Note Some of the values differ slightly (a few % units) from those presented in Table 3 of Hargrave et al (1995). The Hargrave et al. (1995) range values were repeated in (Hargrave et al., 1997).
[2] – Means and standard deviations are from Table 3 in (Bugden et al., 2001).
[3] – Calculated from data in Appendix F of Hargrave et al. (1993a). The values differ slightly from those in Appendix A of the report because Appendix A contains averages of sample means rather than averages of the individual values.

Wildish et al. (2004a) conducted an experiment in which they increased the amount of L-ascorbic acid and SAOB relative to a constant amount of water contained sulfide and measured the sulfide concentration using the ISE method based on a calibration relationship that used a 1:1 ratio of aqueous standard to SAOB. The results indicated that the ISE sulfide estimate of sulfide concentration increased as the amount of SAOB increased. This trend is analogous to the adding of a constant amount of SAOB to sediment with different water contents. In sediments with a high-water content the ratio of SAOB to water is approximately 1:1 whereas in sediments with low water content the ratio of SAOB to water is much greater than 1:1. They suggested the ISE sulfide concentrations should be adjusted for this effect, since without the adjustment the sulfide estimates were negatively biased i.e., the ISE method of keeping a 1:1 SAOB to sediment volume ratio results in an underestimate of the true pore water 'free' sulfide concentration.

Hargrave et al (2008a) recognized this finding and suggested that ISE estimates of concentration of 'free' sulfide in the sediment pore water (S_{pw}) need to account for the fact the

sediment sample is not 100% pore water, it is actually some proportion of the volume of sediment being analyzed. They suggested (Hargrave et al., 2008a) the pore water concentration should be estimated by multiplying the ISE sulfide measurement (S_{ISE}) by the proportion of the volume of the sediment sample that is pore water ($100/WC$) i.e.,

$$S_{pw} = S_{ISE}\ x\ 100/WC.$$

The magnitude of this adjustment can be considerable (Table 3). For sediment with a water content typical of sediments found in aquaculture areas i.e., 50% water content, the adjustment factor is 2.0 resulting in a negative relative bias of 100%. When the water content is accounted for, the sulfide concentration can be reported as µM S-2·mL-1 rather than the more usual µM S-2 per sediment volume, typically 5 mL, that is used when water content considerations are not taken into account (Hargrave et al., 2008b).

Although the magnitude of the adjustment factor can be large, the water content of sediments within the vicinity of a single farm seems to be reasonably consistent, i.e., standard deviations are within approximately 10% of the mean (Table 2). For sediments with a mean water content of 50%, about 95% of the sample water content values would be within 30% to 70% (mean ±2SD) and adjustment factors would range from 3.3 to 1.4 i.e., ± a factor if 1.4 times the mean adjustment factor of 2.

The above relationship suggests that if the concentration of sulfide in the pore water was constant i.e., did not vary between sediment samples with differing water content, the unadjusted ISE estimate of the sulfide concentration would increase with water content ($S_{ISE} = S_{pw}\ x\ WC/100$). Empirical evidence is consistent with this expectation. Reported empirical correlations between ISE sulfides and water content have indeed been positive (Chang et al., 2013). The magnitude of the correlations have varied and this too is to be expected since the concentration of pore water sulfide in sediments from an area, different samples from a site, and/or different samples from a station is not likely to be constant; it is expected to be variable for multiple reasons and the magnitude of this variation will degrade the ISE sulfide to sediment water content correlation. A Pearson correlation coefficient (r2) of +0.667 was reported for S_{ISE} and WC measurements made on sediments collected in association with suspended mussel culture in Tracadie Bay, Prince Edward Island (Hargrave et al., 2008a). Using the same data, the correlation between $log_{10}S_{ISE}$ and WC was 0.764. Pearson correlation coefficients between $log_{10}S_{ISE}$ and WC determined from sediment samples collected in association with marine net pen finfish culture in Nova Scotia were +0.36 and +0.37 (Chang et al., 2013).

The loss of accuracy associated with water content can be mitigated by measuring the water content of sediment samples and making the above adjustment or avoiding taking the extra measurement and extracting the pore water from the sediment and then measuring the sulfide content of the pore water. The measurement of the pore water sample can be made with the ISE method or with other methods. This approach has been adopted in some recent analyses of sediment sulfide (Cranford et al., 2020). The approach takes time to acquire a sufficient amount of pore water for analysis (L. McCrae pers. comm.).

Table 3: The effect of sediment water content adjustments on ISE sulfide measurements. Absolute bias is calculated as $B = S_{ISE} - S_{pw}$ and relative bias is $100\,(S_{ISE} - S_{pw})/S_{pw}$. Negative bias means the initial ISE measurement underestimates the pore water concentration.

Sediment Water Content (%)	AF	Unadjusted ISE Sulfide (S_{ISE} , uM)	Adjusted ISE Pore Water Sulfide (S_{pw} , μM)	Abs Bias (mM)	Rel Bias (%)
10	10.0	1000	10000	-9000	-900
20	5.0	1000	5000	-4000	-400
30	3.3	1000	3333	-2333	-233
40	2.5	1000	2500	-1500	-150
50	2.0	1000	2000	-1000	-100
60	1.7	1000	1667	-667	-67
70	1.4	1000	1429	-429	-43
80	1.3	1000	1250	-250	-25
90	1.1	1000	1111	-111	-11
100	1.0	1000	1000	0	0
10	10.0	3000	30000	-27000	-900
20	5.0	3000	15000	-12000	-400
30	3.3	3000	10000	-7000	-233
40	2.5	3000	7500	-4500	-150
50	2.0	3000	6000	-3000	-100
60	1.7	3000	5000	-2000	-67
70	1.4	3000	4286	-1286	-43
80	1.3	3000	3750	-750	-25
90	1.1	3000	3333	-333	-11
100	1.0	3000	3000	0	0

COLLECTION BY GRAB OR CORE

Sediments collected for ISE determinations of pore water ‘free’ sulfide, are usually collected by grabs or cores. The grabs are deployed from surface vessels and their interaction with the bottom is usually not known on a sample-by-sample basis. Cores can be deployed from surface vessels or divers. Grabs are usually able to obtain a sample over a broader range of sediment grain size than cores; the latter are best used in fine grained sediments.

Video records of some grab sampling activities have shown that at times the bow wave associated with grab sampling when the grab nears the seabed may result in the sediment-water interface being disturbed. Video records of the retrieval of grab samples through

the water column show that at times some flushing or washing of the sediment sample may occur during the retrieval process and that without the video record this may not be detected by looking at the sample once it is on deck. Core samples are generally acknowledged to be associated with less sediment disturbance and hence provide a higher quality sediment sample for analyses. This is particularly true for diver collected sediments. Remotely collected core samples often give high quality samples but video of some core sampling indicates that core samples can also be washed during retrieval if the sampling device does not properly seal the core before retrieval. Unlike samples collected using surface deployed cores and grabs, diver collected samples may be influenced by diver 'bias' in selecting the location of the core.

Given the differences between grab and core sampling it is natural to ask whether there is a difference between ISE pore water 'free' sulfide estimates between sediments collected with grabs and cores. Although, few studies seem to have addressed this question, there have been some comparisons.

Chang et al. (2013) found that the mean sulfide concentration was higher in the grab samples than in core samples collected from 1 of 8 sample stations where concurrent core and grab samples were taken. At each of the 8 stations three grabs and 3 cores were collected. The cores were collected by divers. A single ISE sulfide determination was made on each sample. The mean difference between the mean ISE sulfide derived from grab samples was 50% less than that derived from core samples, although the difference was only marginally significant (p=0.04; Chang et al., 2013). Reasons for the difference were not discussed in Chang et al. (2013). This magnitude of difference is large compared to the uncertainty associated with baseline ISE methodology.

This potential for sampling collection method to result in quite large differences between ISE measurement results needs to be taken into consideration when comparing ISE sulfide values collected from different places and times because differing sampling techniques can potentially account for the differences between measurements. This confounding factor can be mitigated by adopting a standard sampling method that can be utilized in a range of sediment conditions and water depths. Although SCUBA diving may help acquire high quality samples, diving cannot be utilized at all locations due to various safety concerns such as safe diving depths and times as well as regulatory diving safety issues.

AVERAGING OF MEASUREMENTS

The above sections have presented information concerning the uncertainty, bias and precision of some ISE sulfide measurements. In many regulatory situations decisions are made on the mean or average of multiple sulfide measurements taken at a farm. Usually, the mean ($\bar{S}$) is calculated as the sum of the measurements (S_i) divided by the number of measurements (n).

$$\bar{S} = \frac{\sum S_1 + S_2 + \cdots S_i + S_n}{n} \qquad \text{Eq. 9}$$

This calculation gives equal weight to each measurement. This is not the most appropriate way to calculate the mean when the weight associated with each measurement is not the same. In the case of sediment sulfide the uncertainty associated with each measurement is not equal, the uncertainty as indicated by the standard deviation increases with the magnitude of the measurement (Section 2).

When measurements have unequal uncertainties associated with them the best estimate of the mean is the weighted mean (Taylor, 1997). The weighted mean ($\bar{S}_{wav}$) is calculated as the sum

of each weighted measurement and the weights ($w_i = 1/\sigma_i^2$) are calculated as the inverse of the variance associated with the measurement.

$$\bar{S}_{wav} = \frac{\sum_i w_1 S + w_2 S + \cdots + w_n S}{\sum_i w_i} \quad \text{Eq. 10}$$

This weighting gives more weight to the more precise measurements, i.e., those with the smaller variances, and gives less weight to the less precise measurements, those with the larger variances. In the case where the variances (σ_i^2) are equal for all measurements, the weights are equal and the formula reduces to the traditional simple arithmetic mean ($\bar{S}$).

The uncertainty associated with the weighted mean is Eq. 9 (Taylor, 1997); this is also referred to as the standard deviation of the mean or the standard error of the mean. This reduces to the familiar standard error of the mean ($SE = \sigma/\sqrt{n}$) when the weights are equal, i.e., the standard deviations associated with the measurements are equal or constant.

$$\sigma_{wav} = \frac{1}{\sqrt{\sum_i w_i}} \quad \text{Eq. 11}$$

Simulation: Equal Weighting and Bias

To illustrate the difference between unweighted and weighted means for some typical sulfide measurements a simulation was conducted in which ten sulfide measurements were randomly generated from a log-normal distribution with a specified mean and variance (Table 4). The uncertainty, i.e., the precision, associated with each sulfide value has been assumed to be the coefficient of variation (CV) with a value of 5% of the assumed population mean. The standard deviation associated with each measurement has been considered to be constant and was calculated as the sulfide ($\bar{S}$) times the CV in percent ($\sigma = \bar{S} \cdot CV/100$). The value of the assumed CV is a value consistent with the CVs shown in Section 2 for baseline variability in sulfide and the derived standard deviation approximates the values observed for the assumed mean concentration. Although the numbers will change with different inputs and different realizations of the same inputs, the equality of the weightings ensures the difference between the weighted and unweighted means is zero for both the unbiased and biased calculations, and the standard error associated with the constant weighted mean is the same as the standard error for the unweighted mean. When bias is added to the measurements as a constant, i.e., does not vary with the measurement value, the difference between the means of the unbiased and biased measurements is the same as the introduced bias.

This simulation serves to illustrate the effect of constant weights on the means and standard errors as well as demonstrate that the bias in measurements is propagated through the calculations of the means without its value being changed. The next simulation explores the unequal weights and biases.

Table 4: An example of simulated sulfide concentrations and the corresponding summary statistics when measurements are given equal weight and different biases. In (a), the measurements are not biased (bias = 0 µM), whereas in (b) the measurements are biased (bias = -250 µM). In both simulations, ten replicates (n=10) were drawn from a log-normal distribution with a mean of 3.70 µM and standard deviation of 0.025. The unbiased S values are the inverse values of the ten replicates. The biased values are the unbiased values plus the bias, the SD are the standard deviations of the measurement error assumed for each replicate and the weightings are the inverse of the variance (SD^2). In (c), a summary of weighted versus unweighted means in this simulation is presented.

(a)

Simulated S (µM)			**Weights**		**Summary Statistics Unbiased S**		
Unbiased S	**Bias**	**Biased S**	**SD**	**Weights**	**Outputs**	**Values**	**Units**
5094	0	5094	250	0.00002	n	10	N/A
5107	0	5107	250	0.00002	a Unweighted Mean	5073	µM
4856	0	4856	250	0.00002	b Weighted Mean	5073	µM
4824	0	4824	250	0.00002	Difference	0	µM
4720	0	4720	250	0.00002	SD with a	265	µM
5381	0	5381	250	0.00002	SD with b	265	µM
5527	0	5527	250	0.00002	CV with a	5.2	%
5312	0	5312	250	0.00002	CV with b	5.2	%
5021	0	5021	250	0.00002	SE Unweighted	84	µM
4885	0	4885	250	0.00002	SE Weighted	84	µM

(b)

Simulated S (µM)			**Weights**		**Summary Statistics Biased S**		
Unbiased S	**Bias**	**Biased S**	**SD**	**Weight**	**Outputs**	**Values**	**Units**
5094	-250	4844	250	0.00002	n	10	N/A
5107	-250	4857	250	0.00002	a Unweighted Mean	4823	µM
4856	-250	4606	250	0.00002	b Weighted Mean	4823	µM
4824	-250	4574	250	0.00002	Difference	0	µM
4720	-250	4470	250	0.00002	SD with a	265	µM
5381	-250	5131	250	0.00002	SD with b	265	µM
5527	-250	5277	250	0.00002	CV with a	5.5	%
5312	-250	5062	250	0.00002	CV with b	5.5	%
5021	-250	4771	250	0.00002	SE Unweighted	84	µM
4885	-250	4635	250	0.00002	SE Weighted	84	µM

(c)

Summary of Means	Unbiased	Biased	Difference (µM)	Difference (%)
Unweighted Mean	5073	4823	-250	-5
Weighted Mean	5073	4823	-250	-5

Simulation: Unequal Weighting and Bias

To illustrate the effect of unequal weights and biases on the difference between unweighted and weighted means the same scenario as above was simulated again but with unequal weights (Table 5). Once again, the sulfide values were randomly generated from a log-normal distribution with same log mean and variance of the log distribution as above. The precision associated with each sulfide value has been assumed to be the coefficient of variation (CV) with the same value of 5% of the mean of the generated sulfide values. The standard deviation associated with each measurement has been calculated as the magnitude of each sulfide measurement (S) times a constant CV in percent ($\sigma = S \cdot CV/100$). This generates variable standard deviations that approximate the observed trend of the precision in measurements decreasing with the magnitude of the sulfide i.e., the standard deviation increases with the magnitude of the sulfide while the CV remains constant or independent of the sulfide magnitude. The weighted and unweighted means and their respective standard errors differ under these assumptions; the difference between the means is about 1% with the weighted mean being less than the unweighted mean. This is as expected because more weight is given to the more precise measurements. As with the equal weight simulation, the bias added to measurements propagates through the calculation such that the difference between the mean of the unbiased and biased measurements equals the magnitude of the introduced bias.

The differences between the unweighted and weighted means increases substantially when the variability in the measurements increases (Table 6). This is illustrated by a third simulation in which the variation in sulfide measurements is high and in accordance with that reported by Cranford, et al. (2011) for ISE measurements made in association with mussel farming. In this case the unweighted and weighted means differ by over 200%. This large difference occurs because the weighted average gives little weight to the large sulfide values and the unweighted average gives equal weight to these values. The difference is also large enough that a different regulatory decision would result since several thresholds are exceeded by the unweighted mean that are not exceeded by the weighted mean. As in the other cases, the bias propagates through the calculations such that the difference between unbiased and biased means is equal to the assumed bias.

Table 5: An example of simulated sulfide concentrations with low variability and the corresponding summary statistics when measurements are given equal (a) and unequal weights (b). In (a), the measurements are not biased (bias = 0 µM), whereas in (b), the measurements are biased by -5%. In both simulations, ten replicates (n=10) were drawn from a log-normal distribution with a mean of 3.70 µM and standard deviation of 0.025. The unbiased S values are the inverse values of the ten replicates. The biased values are the unbiased values plus the bias, the SDs are the standard deviations of the measurement error (5%) assumed for each replicate and the weightings are the inverse of the variance (SD^2). In (c), a summary of the means in this simulation is presented.

(a)

Simulated S (µM)			Weights		Summary Statistics Unbiased S		
Unbiased S	**Bias**	**Biased S**	**SD**	**Weight**	**Outputs**	**Values**	**Units**
4721	0	4721	236	0.00002	n	10	N/A
5088	0	5088	254	0.00002	a Unweighted Mean	4847	µM
5043	0	5043	252	0.00002	b Weighted Mean	4816	µM
4435	0	4435	222	0.00002	Difference	31	µM
5423	0	5423	271	0.00001	SD with a	282	µM
4626	0	4626	231	0.00002	SD with b	284	µM
4818	0	4818	241	0.00002	CV with a	6	%
4865	0	4865	243	0.00002	CV with b	6	%
4827	0	4827	241	0.00002	SE Unweighted	89	µM
4620	0	4620	231	0.00002	SE Weighted	90	µM

(b)

Simulated S (µM)			Weights		Summary Statistics Biased S		
Unbiased S	**Bias**	**Biased S**	**SD**	**Weights**	**Outputs**	**Values**	**Units**
4721	-236	4485	236	0.00002	n	10	N/A
5088	-254	4834	254	0.00002	a Unweighted Mean	4604	µM
5043	-252	4791	252	0.00002	b Weighted Mean	4577	µM
4435	-222	4213	222	0.00002	Difference	28	µM
5423	-271	5152	271	0.00001	SD with a	268	µM
4626	-231	4395	231	0.00002	SD with b	270	µM
4818	-241	4577	241	0.00002	CV with a	6	%
4865	-243	4621	243	0.00002	CV with b	6	%
4827	-241	4586	241	0.00002	SE Unweighted	85	µM
4620	-231	4389	231	0.00002	SE Weighted	85	µM

(c)

Summary of Means	Unbiased	Biased	Difference (μM)	Difference (%)
Unweighted Mean	4847	4604	-242	-5
Weighted Mean	4818	4577	-241	-5
Difference in Means (μM)	28	27	-	-
Difference in Means (%)	1	1	-	-

Table 6: An example of simulated sulfide concentrations with high variation and the corresponding summary statistics when measurements are given larger unequal weights. In (a), the measurements are not biased (bias = 0 μM), whereas in (b), the measurements are biased by -5%. In both simulations ten replicates (n=10) were drawn from a log-normal distribution with a mean of 3.70 μM and standard deviation of 0.5. The unbiased S values are the inverse values of the ten replicates. The biased values are the unbiased values plus the bias, the SDs are the standard deviations of the measurement error (5%) assumed for each replicate and the weightings are the inverse of the variance (SD^2). In (c), a summary of the means in this simulation is presented.

(a)

Simulated S (μM)			Weights		Summary Statistics Unbiased S		
Unbiased S	Bias	Biased S	SD	Weight	Outputs	Values	Units
9074	0	9074	454	0.00000	n	10	N/A
16314	0	16314	816	0.00000	a Unweighted Mean	6598	μM
8373	0	8373	419	0.00001	b Weighted Mean	2126	μM
1299	0	1299	65	0.00024	Difference	4472	μM
8200	0	8200	410	0.00001	SD with a	8261	μM
3995	0	3995	200	0.00003	SD with b	4475	μM
8270	0	8270	414	0.00001	CV with a	125	%
5085	0	5085	254	0.00002	CV with b	210	%
3992	0	3992	200	0.00003	SE Unweighted	2612	μM
1376	0	1376	69	0.00021	SE Weighted	1415	μM

(b)

Simulated S (μM)			Weights		Summary Statistics Biased S		
unbiased S	Bias	Biased S	SD	Weights	Outputs	Values	Units
9074	-454	8620	454	0.00000	n	10	N/A
16314	-816	15498	816	0.00000	a Unweighted Mean	6268	μM
8373	-419	7954	419	0.00001	b Weighted Mean	1927	μM
1299	-65	1234	65	0.00024	Difference	4249	μM
8200	-410	7790	410	0.00001	SD with a	7848	μM
3995	-200	3796	200	0.00003	SD with b	4251	μM
8270	-414	7857	414	0.00001	CV with a	125	%
5085	-254	4831	254	0.00002	CV with b	211	%
3992	-200	3792	200	0.00003	SE Unweighted	2482	μM

Simulated S (μM)			Weights		Summary Statistics Biased S		
unbiased S	Bias	Biased S	SD	Weights	Outputs	Values	Units
1376	-69	1307	69	0.00021	SE Weighted	1344	μM

(c)

Summary of Means	Unbiased	Biased	Difference (μM)	Difference (%)
Unweighted Mean	6598	6268	-330	-5
Weighted Mean	2028	1927	-101	-5
Difference in Means (μM)	4569	4341	-	-
Difference in Means (%)	225	225	-	-

Simulation: Logarithmic Transformation of Sulfide Measurements

The frequency distribution of ISE sulfide values is right skewed (the distribution has a prolonged tail to the right of the mode) and the variance in ISE sulfide measurements increases with the mean concentration and the coefficient of variation is approximately independent of the mean concentration, consistent with the standard deviation increasing with the magnitude of the measurement.

Logarithmic transformation of sulfide values shifts the distribution of values toward a more symmetric log-normal distribution, i.e., a normal distribution of logarithmic values, and stabilizes the variance i.e., makes the variance independent of the mean. These effects are illustrated in Figure 11. The transformation does not affect the ranking of measurements and does not influence non-parametric hypothesis tests whose probability calculations depend on symmetry around the sample median.

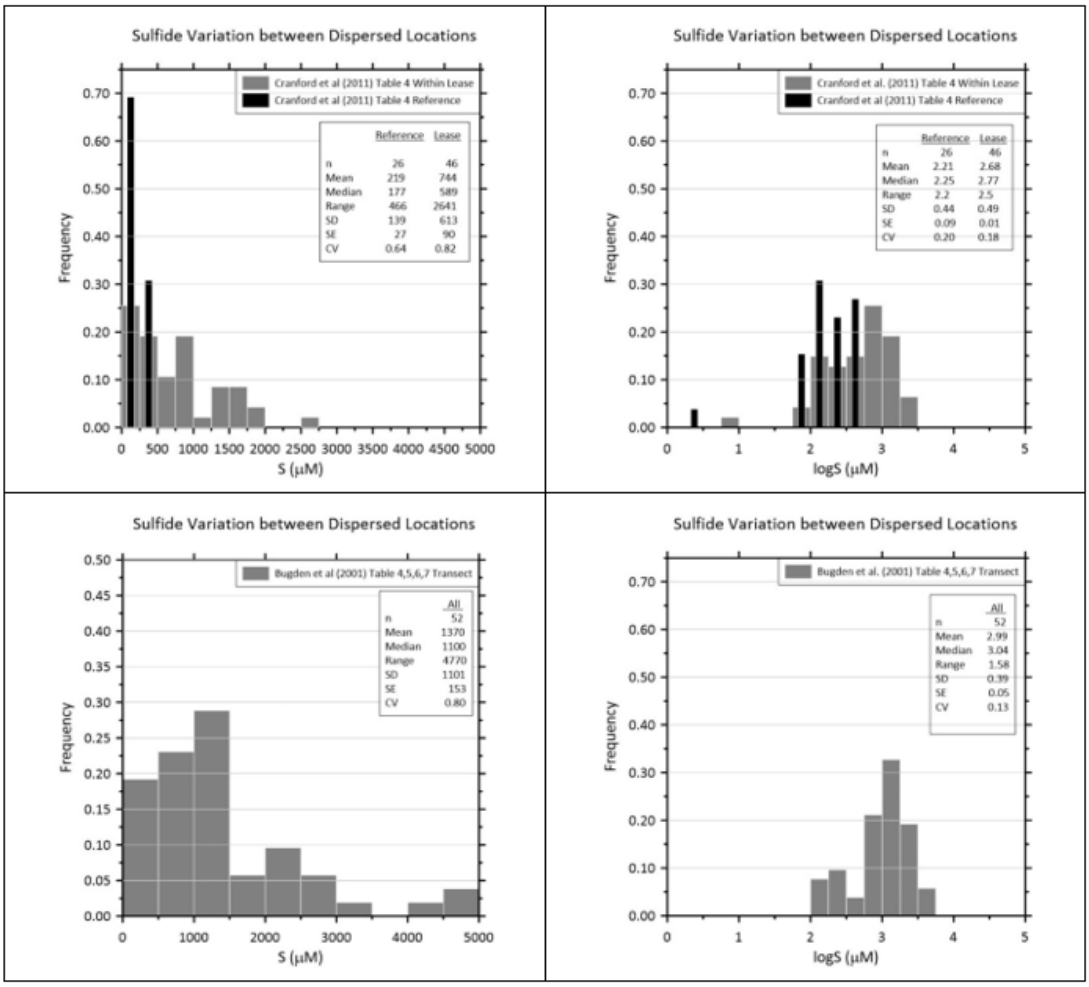

Figure 11: Examples of the effect of a logarithmic transformation of ISE sulfide concentration estimates on the frequency distribution of the data. The left panels show untransformed and the right panels show log_{10} transformed ISE sulfide measurements. The data in the top panels are reported in Cranford, et al. (2011) and is from a coastal marine shellfish farming area. The data in the lower panels are from (Bugden et al., 2001) and is from a coastal finfish farming area. Both sites are in Atlantic Canada.

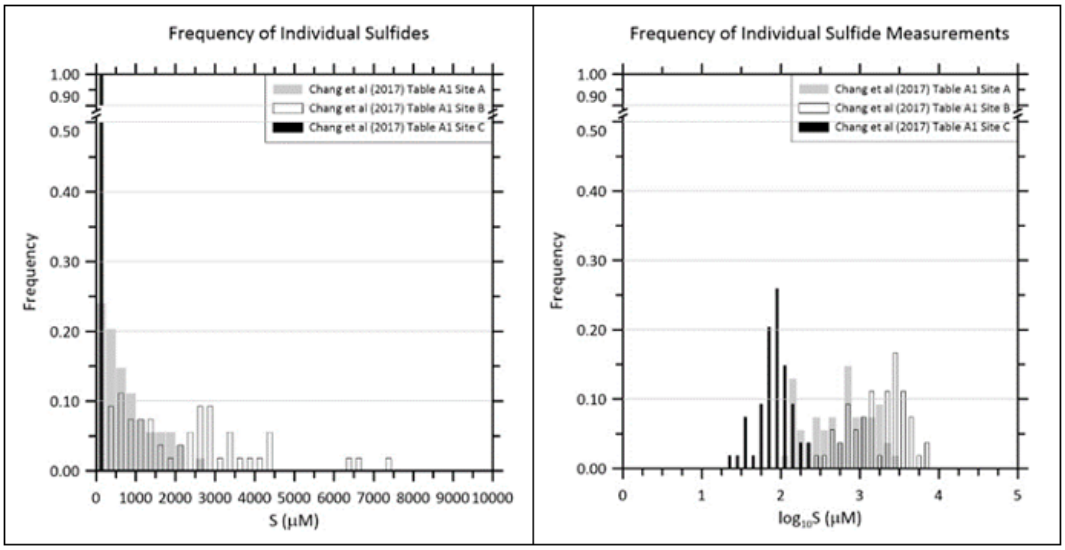

Figure 11 continued: The data in theses panels are reported in Chang et al. (2017) and are from a coastal marine finfish farming area in Atlantic Canada.

SECTION 4: IMPLICATIONS TO ENVIRONMENTAL MONITORING

CONTEXT: MAGNITUDE OF ISE SULFIDE IN RELATION TO TEMPORAL VARIATION

A distinct seasonality with an annual maximum in September was found in measurements taken in the Fundy Isles region of the Bay of Fundy (Hargrave et al., 1993b; 1997). Another study that was conducted near an open net pen salmon farm located in the Passamaquoddy Bay area of the southwest New Brunswick portion of the Bay of Fundy, and a few kilometers north of the Fundy Isles region, also indicated a maximum in sediment sulfide in September and October (Wildish et al., 2005). In contrast, monthly sampling at two different locations within the Fundy Isles region, one location on the perimeter of a salmon farm net pen array and the other several hundred meters away from the farm and near a tidally inundated bar, found considerable variation in ISE sulfide determinations but no evidence of seasonality or an annual maximum in September (Wildish, Akagi, & Martin, 2002). This finding was similar to that found by a more temporally restricted study in the Lime Kiln and Bliss Harbour area of the Fundy Isles conducted by Page et al. (2010). Sampling in this study was conducted over the July – September period, and results indicated considerable temporal variation but only a weak temporal trend (Page et al., 2010). The evidence does not therefore indicate a clear and consistent seasonality. Temporal patterns seem to vary between sites and years, and detection of spatial and temporal patterns are confounded by measurement variability as well as in situ spatial and temporal variability. Hence, there may be value in giving more consideration to the value of an improved statistical quantification of the uncertainty associated with regulatory indicators i.e., what level of uncertainty is deemed acceptable to support regulatory needs and what level of sampling is needed to satisfy these needs. Insights from these considerations could be used to help determine if more information is needed, and if so, help guide the development of more comprehensive research sampling and analyses that could lead to additional advice related to regulatory oriented sampling. For example, higher temporal resolution of sampling would contribute to a better quantification of the statistical character of temporal variability which in turn could contribute to determining whether the timing of regulatory sampling influences results

and if so help determine when regulatory sampling should be conducted to achieve the regulatory goals.

Although the magnitude of the natural temporal variation in ISE sulfide concentrations is important when interpreting ISE sulfides measurements in relation to thresholds it is not the focus of this document; this document is focused on the factors influencing the uncertainty of ISE measurements and their averages when samples have been collected at a single point in time or within a narrow window (hours) of time. Certainly, if samples are collected over a wide time window, temporal trends in sulfide concentrations may influence estimates of variation between samples and sample means.

CONTEXT: MAGNITUDE OF ISE SULFIDE VARIATION IN RELATION TO SPATIAL SCALE

The magnitude of variability in ISE sulfide measurements has implications for the consistency, and hence credibility, of regulatory results as well as how regulatory summary statistics should be calculated. This point has been made several times by various authors in the context of both finfish (Chang et al., 2011) and shellfish (Cranford et al., 2011) monitoring programs.

In this section the variability in the baseline ISE sulfide measurements is placed into the context of the variability in ISE measurements associated with sampling sediments collected from the marine coastal zone in which marine aquaculture takes place. This has been done by grouping estimates of the coefficient of variation in ISE sulfide measurements into categories representing the approximate distance between measurements and ordering the categories from the least to most variable. The categories and the sources of information associated with each are listed and described in Table 7 and the distribution of coefficients of variation (CV) associated with each category is shown in Figure 12.

Precision

The CVs and the range in CV values increases as the distance between the samples of subsamples increases Figure 9. The smallest CVs are associated with the laboratory measurements on fully homogeneous standard sulfide solutions and the largest CVs are associated with reference areas and aquaculture farm sites. The CVs associated with the measurements on standard solutions represents the baseline variability; the median CV is approximately 0.05 (5%), the minimum is less than 0.02 (2%) and the maximum is less about 0.1 (10%). All other categories involve measurements made on sediment. The median CV associated with measurements made on homogenized sediment is similar to that made on the standard solutions i.e., 0.04 (5%) but the spread of values is larger than that for standard solutions with the minimum being less than 0.01 (1%) and the maximum CV being about 0.3 (30%). These patterns are presumably associated with the more complex nature of the homogenized sediment relative to the solution.

The variation in subsamples taken from a single sampling unit such as a grab, is greater than observed for homogenized samples because of spatial variation within the grab. The median CVs associated with reference and farm sites are respectively 0.6 (60%) and 0.7 (70%). This is an order of magnitude greater than the median CVs for standard solutions and homogenized sediments. The increase in the maximum CVs is even greater whereas the increase in the minimum CVs is less than the increase associated with the median, i.e., the range in CVs associated with reference and farm sites is also an order of magnitude larger than for measurements on standard solutions.

Bias

The above comparisons do not consider bias in ISE measurements. Bias is difficult to determine for field measurements since a true value or very precise estimate of in-situ sulfide concentration is generally not available. Unfortunately, the introduction of bias can lead to false conclusions when field measurements are compared to threshold values and this cannot be compensated for by changes to sample number; it can only be minimized by the use of good standard methodology.

In general, bias in measurements does not affect the precision of the measurements when the bias is equal for all measurements. This is the case when laboratory procedures introduce a bias across all the measurements associated with a particular sample batch. However, if the bias varies between measurements the bias becomes part of the variability in the measurements. For example, failing to account for sediment water content certainly has the potential to generate considerable bias in ISE sulfide measurements, however, it will only affect variation in ISE sulfide measurements when the water content is not constant, which is usually the case, and the magnitude of the effect is proportional to the variability in water content between sediment samples. As indicated in Table 2 (included in the Sediment Water Content section of this report) the mean water content of sediments is typically about 50% with a standard deviation of about 10%. The corresponding coefficient of variation in water content is therefore of order 0.2 (20%). This value can often be more in the order of 10%. This variability is already included in the scale of reference and farm site variation shown in Figure 12 since the ISE measurements in the studies included in the generation of the figure did not correct for water content.

Table 7: Description of data sources and sample groupings used for estimation of the ISE sulfide variation shown in Figure 9. In the table 0() denotes in the order of.

Scale	Description	Source(s)
ISE Standard	Measurements made on sulfide standard aqueous solutions that have not been stored and measurements have been made with freshly calibrated electrodes using titrated stock solutions. The solutions have no sediment. Measurements are separated by distances of effectively O(0.0 m).	(Chang et al., 2014)
Mixed Sediment	Measurements made on homogenized sediment. Measurements are separated by distances of effectively O(0.0 m).	(Wildish et al., 1999; 2004); (Wong and Page, 2025b)
Within Samples	Measurements made within the confines of a sampling unit such as a grab. Measurements are separated by distances of O(0.1 m).	(Chang et al., 2011; 2013; 2017); (Wildish et al., 2002; 2004)
Within Stations	Measurements made at the sampling station. Multiple grabs or cores were collected from the same surface location.	(Chang et al., 2011; 2013; 2017)

Scale	Description	Source(s)
	Measurements are separated by distances of O(10 m).	
Within Ref. Areas	Measurements made at multiple stations within are separated by distances similar to the length scales of fish farms, Measurements are separated by distances of O(100 m).	(Chang et al., 2017); (Cranford et al., 2011); (Bugden et al., 2001)
Within Farm Sites	Measurements made at multiple stations within and around a fish farm. Measurements are separated by distances of O(100 m).	(Chang et al., 2017); (Cranford et al., 2011); (Hargrave et al., 1998)

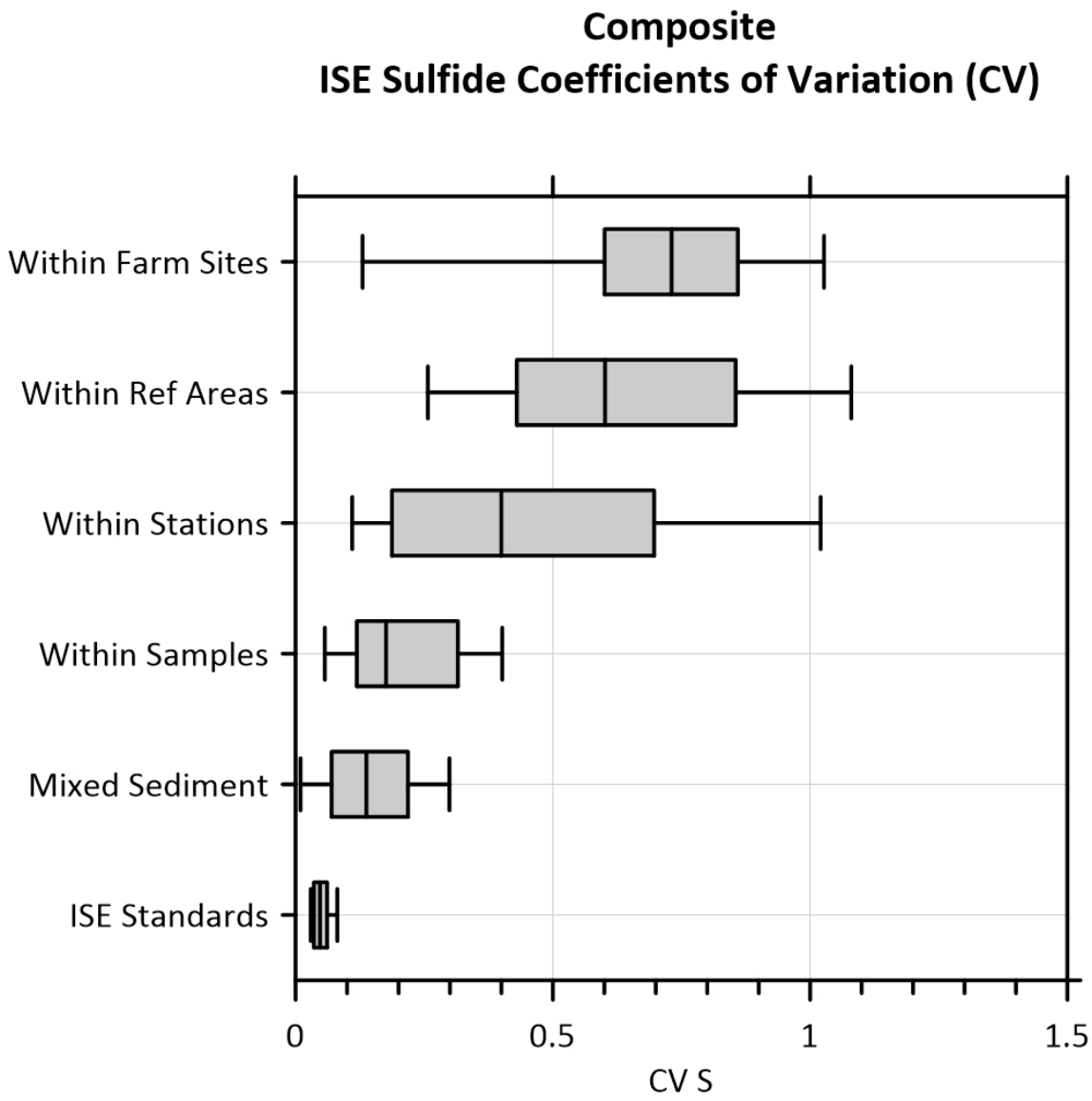

Figure 12: The variation in ISE sulfide measurements in relation to the separation distance between samples. The whisker extremes represent the 10% and 90% values, the shaded box limits represent the 25th and 75th percentiles of the values and the central line represents the median value. The sources of data are listed in Table 7.

IMPLICATIONS OF FACTORS AFFECTING ISE SULFIDE MEASUREMENTS

The pattern of increasing measurement variation with distance between ISE measurements and the simulations related to the calculation of averages indicates that increases in the precision of ISE measurements is not likely to reduce the variation in environmental sampling or monitoring results and site means. This is because the largest contributor to the variation in environmental results is spatial variation in the environment rather than sulfide measurement methodology, especially when the ISE measurements adhere to the ISE measurement practices described in Wong and Page (2025a) such as timely and frequent calibration, limited storage of standards, electrodes, and samples, and accurate titration of standard stock solutions. Improvements to the robustness (repeatability) of environmental monitoring decisions will therefore need to be largely based on reducing the environmentally induced variation in ISE measurements.

ISE METHOD CONSIDERATIONS

Electrode drift

Electrode drift can increase the uncertainty or reduce the precision of ISE measurements by about 5% over a twelve-hour period (Section 1) and introduce a negative bias of over 10% after this time period (Section 1). In other words, if electrode drift was the only factor increasing the baseline uncertainty the CV would increase from a baseline of approximately 5% to between 5% and 10% (Section 3). This is still small in relation to the variation associated with environmental measurements.

The effect of electrode drift can be mitigated by ensuring the electrodes are frequently calibrated and drift has not been detected. Although a post-calibration use time of less than 3 or 4 hours is often adopted, this does not preclude all drift since the probe history and exposure to poisoning conditions may accelerate the drift. Robust criteria for this check need to be developed so the magnitude of drift can be kept below a specified amount.

Storage of Calibration Standards

The sulfide concentration in calibration standards decreases with storage time resulting in a negative bias in ISE estimates of sulfide concentration (Section 3). The magnitude of the bias decreases with increasing sulfide nominal concentration. The magnitude of the bias typically exceeds 10% and can be more than 70%. The precision of the ISE measurements decreases with storage time, i.e., the variability in measurements increases with storage time. The magnitude of the variability increases from less than 10% to several tens of percent (Section 3).

When calibration standards are stored the magnitude of the bias and precision can be similar to the magnitude of the variability in the environment. This is not a preferred situation since the magnitude of variance in ISE measurements of field samples should primarily reflect the in-situ magnitudes and variance of the sulfides.

The finding that bias changes with storage of sulfide standard solutions suggests that storage of pore water extracted from sediments will also result in bias that is larger than that associated with no storage. The UV method developed by Cranford et al. (2020) and the methylene blue method developed by Wong and mentioned by Cranford et al. (2020) avoids some of the sediment matrix induced bias by extracting the pore water as soon as the sample is collected. The UV method can further minimize bias by taking the measurement on the pore water immediately, i.e., within a few minutes, following extraction and the methylene blue method can reduce bias by adding fixative to the extracted pore water immediately after extraction. Both of these practices have been implemented when the methods have been used by the developers of the methods (Cranford et al., 2020).

Fortunately, both of these effects can be mitigated by ensuring standard solutions are freshly made prior to each calibration session and certainly not stored for more than a few hours i.e., only used during the day they were made.

Storage of Samples

As in the storage of sulfide standard solutions, the sulfide concentration in sediment changes with storage time (Section 3; Wong and Page, 2025b). The storage introduces a negative and variable estimated bias into the measurements and their precision remains independent of storage time (Section 3).

The bias associated with an environmental sample is generally not known, since it can only be estimated by taking measurements at various times of storage including a measurement taken very soon after the sample has been collected. When there is virtually no storage the bias

appears to be about ±5%, at least as indicated by the results from experiments conducted by a single laboratory. The evidence examined suggests both the sign and magnitude of the bias is variable but is often about 20% although it can sometimes exceed 40% (Section 3; Wong and Page 2025b). At very short storage times the bias is small in relation to the variation associated with most environmental measurements. However, it can be comparable to the environmental variability in some cases.

The precision of the ISE measurements appears to be variable, centered around 10% with some CVs exceeding 20%; the CVs are independent of storage time (Section 3). These CVs are probably an underestimate since they are based on homogenized sediment.

The storage effects can be mitigated by limiting the storage time of samples and ideally making the ISE measurements within minutes of sediment collection (bias begins to appear very quickly).

Adjustment for Sediment Water Content

Adjustment for water content can make substantial differences in sulfide values (Section 3). When the samples are collected from an area in which the sediment water content is homogeneous with a low CV, not often the case, the water correction will not influence the CV of ISE measurements. However, if the sediment water content is not homogeneous, the usual case, it can contribute substantially to the variability in ISE sulfide measurements at the reference and site scales.

The effect of water content can be mitigated by measuring the water content for each sample and appropriately adjusting the ISE sulfide values or by extracting the pore water from the sediment and analyzing the pore water for sulfide (Hargrave et al., 2008a; Wildish et al., 2002). Rhizomes have been used for this purpose in association with the recent exploration of UV and methylene blue methods of measuring sediment sulfide (Cranford et al., 2020). At least a couple of milliliters of pore water would be required for an ISE analyses of the pore water sulfide since the electrode needs to be sufficiently immersed in the pore water. Further work is needed to assess the degree to which pore water extraction reduces variation in sulfide measurements.

HYPOTHESIS TESTING: STATISTICAL CONSIDERATIONS

The desire of many regulators is to keep environmental impacts below specified thresholds. In the case of Canada, the conceptual desire is to keep reductions in biodiversity to specified levels, and since biodiversity analyses are time consuming and expensive (Wildish et al. 2001; 2004b), the concentration of free sulfide in surficial sediments has been adopted as a proxy or indicator of declines in biodiversity and the regulatory desire is to keep free sulfide concentrations in the sediment below specified thresholds (Wildish et al., 1999; Hargrave et al., 2008b, Chang et al., 2011, AARs)[1].

In statistical terms, the regulatory objective is to test the null hypothesis (Eq. 10) that the mean (or median) measured free sulfide concentration ($\overline{S}_M$) is consistent with the background mean (or median) concentration ($\overline{S}_B$).

[1] Note: It is well recognized that free sulfide is not a good indicator of biodiversity recovery, because biodiversity can only occur once sulfide concentrations have recovered allowing biological recolonization and stabilization to occur. The sulfide recovery time scale is often shorter than the biodiversity recovery time scale.

$$H_0: \bar{S}_M = \bar{S}_B \quad \text{Eq. 12}$$

The alternative hypothesis (Eq. 11) is that the measured free sulfide concentration is greater than some specified sulfide mean (or median).

$$H_0: \bar{S}_M > \bar{S}_B \quad \text{Eq. 13}$$

Tests of statistical significance are often sensitive to the assumption of the data being normally distributed (van Belle, 2002; Qian, 2010) but the tests are often more sensitive to the assumption of data value independence (Qian, 2010).

In order to be able to test the above hypotheses, several things must be determined or assumed. These include: the uncertainty associated with each measurement, the variability associated with the in-situ measurements, the degree of difference between the background and the "impacted ", i.e., measured, sulfides and the likelihood that the empirical difference actually exists (alpha level and probability of a Type I error).

In the Canadian regulatory environment, the sulfide impact thresholds have been set (Appendix 3). The sample size and sampling designs have also been set and they vary among jurisdictions. Although a detailed consideration of the effect of ISE bias and precision on sampling designs is beyond the scope of this document, a few preliminary comments are made in the following sections. A more thorough consideration would be worthwhile undertaking at a later date.

Sample Size and Bias Effects

One specific category of consideration is sampling design (sample number, spatial allocation of samples). Although an increase in sample number may not substantially affect the magnitude of ISE variability, it will reduce the standard error of the mean and hence improve the power of statistical tests. An optimal sampling design may reduce the variation and this may also increase the ability to detect the differences of interest to regulators. Regulators are often most interested in the period of the year when maximum sulfide values are expected, so potential temporal variation in ISE measurements can be mitigated by ensuring sampling is conducted during the appropriate window defining maximum sulfide concentrations.

No matter when a sampling design is executed one of the first questions that is often asked is how many samples should be taken to detect the desired effect. A simple rule of thumb formula (Eq. 10) estimates the number of samples (n) as a function of the expected coefficient of variability (CV) in the measurements and the proportionate change (PC) to be detected (van Belle, 2002). In this calculation the PC is calculated as the difference between means divided by the initial value.

$$n = \frac{8(CV)^2}{(PC)^2}[1 + (1 - PC)^2] \quad \text{Eq. 14}$$

Use of this equation for values of CV and PC that may be representative of some aquaculture impact considerations indicates that sample sizes can be fairly small, i.e., about 5, for many situations. For example, if the regulator wishes to detect whether a background mean sulfide concentration of 1,000 μM has changed to 1,500 μM the PC is (1,000-1,500)/1,000 which equals 0.5. The number of samples needed to detect this change varies with the CV and is

small (<4) when the CV is 0.3 or less. However, when the CV is 0.5 or greater, the number of samples needed is greater than 10, it is 23 for a typical farm site CV of 0.75 and is 40 when the CV is 1. When the proportional change increases to a value of 1.0 as is the case when the interest is in whether a farm site mean sulfide is 1,500 μM or 3,000 and the regulator wishes to know if the mean has increased to 3,000 μM or 6,000, the sample size for a typical CV of field samples of 0.75 is 5.

These calculations illustrate that the ability to detect differences between site means and thresholds is not overly restricted by the need for large sample sizes, in fact relatively small sample sizes seem to be sufficient for many practical situations.

In the situation where multiple samples are taken at each station, such as required by the Nova Scotia Environmental Monitoring Program (NSFA, 2011), an average of multiple sulfide measurements should help to reduce the variability at the within a station scale. The approach of setting thresholds on the proportion of station means that fall above or below a threshold used by Nova Scotia (NSFA, 2011) is interesting and its effectiveness in stabilizing site classifications should be examined from the perspective of the location and distance between stations relative to the magnitude of sulfide variability at the farm and between station scales.

As was indicated earlier, it is the potential for bias in sulfide measurements that should be of most concern and increases to sample size will not reduce systematic bias. As indicated in Section 2 if sediment samples are stored the bias can routinely be 20% or more. Assuming the relatively small number of samples collected from a site are analysed as a single batch any bias associated with the sample collection, storage and analyses will have a direct impact on the site mean. For example, a 20% bias, a value easily generated by sample storage will result in a site mean of 3,000 μM to values of 2,400 μM (negative bias) or 3,600 μM (positive bias) i.e., the difference that brackets a regulator classification jump. The corresponding values for a bias of 0.4 (40%) are 1,800 μM to 4,200 μM. Since the biases associated with any specific sampling and analyses implementation are not known and cannot be mitigated by increasing the sample size, it seems prudent to mitigate the potential for this magnitude of bias by not storing samples, electrodes, standard solutions etc.

Wilcoxon Signed Rank Test

The Wilcoxon Signed Rank Test is a nonparametric test and compares the median of a single set of values, to a specified value (Sheskin, 2007). In the context considered here, the values are sulfide concentration measurements (S) and the constant specified value is a regulatory threshold (S_{reg}).

The assumptions of the test include (Sheskin, 2007):

The measurements are continuous.

The measurements were independent and randomly selected.

The distribution of the measurements are from a distribution that is symmetrical about its median, i.e., there are an equal number of measurement values above and below the median.

The distribution of samples can be Gaussian or Normal but does need to be. However, if the distribution is Normal, a parametric test is more powerful than the Wilcoxon test (Sheskin, 2007).

When the values are from a symmetrical distribution the null hypothesis is that the median of the sulfide values (S_{md}) is equal to the specified value (Sheskin, 2007), i.e.,

$$H0: S_{md} = S_{reg}.$$

When the measurements are from a distribution that is not symmetrical the hypothesis is that the measurements are different (or not) but not that the medians are different (Sheskin, 2007).

The alternative hypotheses are:

$$HA: S_{md} \neq S_{reg}$$

$$HA: S_{md} < S_{reg}$$

$$HA: S_{md} > S_{reg}$$

The HA of interest in the regulatory context is that the median of the measurements is greater than (>) the regulatory threshold.

The test statistic generated by the Wilcoxon test is T (described below) and the two-sided *p* value associated with the test statistic indicates the probability that the sample values have come from a symmetrical population with a median that equals the specified value. When the alterative hypothesis of interest is that the samples have come from a distribution that has a median greater than the constant value, a one-sided *p* value is used (Sheskin, 2007).

The test generates a p-value which is the chance of seeing the result assuming the null hypothesis is true i.e., the measurements come from a population that includes the threshold value or a population that has the same median.

As with all statistical tests the probabilities that the test statistic will have a value equal to a particular value are dependent on the sample size and the distribution of the sample values.

In the case of the Wilcoxon test, the test will always reject the null hypothesis (Ho) the sample median is greater than the specified threshold if no measurement value is greater than the constant value, i.e., the measurements do not bracket the regulatory threshold i.e., if the threshold does not fall within the range of the measurements. Also, the test will always give a *p* value greater than 0.05 when $n \leq 5$ (Appendix 7). To get a p value smaller than 0.05 the sample size needs to be larger or the number of measurements exceeding the threshold needs to increase.

SECTION 5: DISCUSSION

One of the primary goals or objectives of aquaculture environmental impact regulations is to limit the degree of reduction in benthic biodiversity. The biodiversity is related to the free sulfide concentration within the pore water of sediments (Hargrave et al., 2008b) and since measuring sediment sulfide is quicker, easier, and cheaper than measuring biodiversity, free sulfide has been chosen in Canada, as an indicator of soft bottom organic impact (DFO, 2018).

The ISE method of measuring pore water sulfide was first introduced in the later 1990s and early 2000s (Wildish et al. 1999; 2004a). The method has come under some criticism over the years for various reasons, including lack of repeatability at monitoring sites which leads to uncertainty in the merit of the method.

The considerations in this document attempt to place the uncertainty associated with ISE methods and some of the major factors influencing its variability into the context of magnitude of variability associated with samples taken from the natural (reference sites) and impacted environment (aquaculture farm sites). Although the document alludes to the fact that bias and

precision has an effect on hypothesis testing and the determination of whether thresholds are exceeded, it does not delve extensively into the details of these statistical implications. This could be the topic of a separate document.

The document suggests that the precision associated with the ISE measurement methodology, when equipment is properly and freshly calibrated and samples are not stored, is typically about an order of magnitude less than the variability between samples collected at various locations within a farm or reference site. However, in some circumstances when the site variability is low, the methodological variability can be similar to the site variability. The bias associated with properly and freshly calibrated electrodes and samples that are not stored is low but increases to tens of percent when storage issues come into play. This highlights the importance of conducting rigorous ISE measurements in an effort to keep measurement bias and uncertainty as low as the method and practical considerations will allow.

BASELINE VARIABILITY IN ISE SULFIDE

The Canadian Federal Aquaculture Monitoring Standard associated with the Canadian Aquaculture Activity Regulations specifies that a millivolt (mV) meter with a 0.1 mV resolution and ion electrode selective probe with an accuracy of ± 5 % is desired and that “the concentration of free sulfide in the sediment samples must be determined within 36 hours of collection, except in British Columbia where sediment samples must be analyzed within 5 minutes” (DFO, 2018). The meter resolution of 0.1 mV should be accompanied by a statement of meter accuracy or trueness since resolution does not imply accuracy.

The information considered in this document indicates the 5% accuracy criteria best corresponds to a central tendency for baseline bias, precision and root mean square error (RMSE) of the ISE method under ideal conditions (titrated stock solutions, freshly made standards, measurements made with freshly calibrated electrodes and meters, no sediment matrix effects, experienced analysts, etc.). When these ideal or baseline conditions are not met, the bias, precision and RMSE can frequently exceed 5% and be 10% or more.

FACTORS DEGRADING ACCURACY

Several factors were examined in terms of their effect on ISE bias and precision. As presented in Wong and Page (2025b), storage of calibration standards and samples introduces bias in sulfide measurements, assuming the measurement of interest is the in situ undisturbed sulfide concentration which is best indicated by the measurements taken as soon as the sediment sample is collected. The magnitude of storage bias differed between sulfide concentrations, sediment matrices and laboratories and varied from a few percent to over 80%. The smallest bias was consistently associated with solutions that had not been stored and the maximum magnitude of this bias was 6%. It would therefore seem prudent that ISE standard operating procedures, if they are intended to minimize bias, should specify that standard solutions should not be stored.

Table 8 summarizes the sources of error associated with the analytical process and contrasts the recommended procedure with procedures that introduce error. The sources are listed according to the magnitudes of error. The major sources of bias, variation and uncertainty in the analytical process are associated with pore water content, storage of standards and electrodes followed by variation or lack of stability in mV readings. There are other sources contributing to analytical variability that have not been quantified in this assessment such as deoxygenation of the water used to make up the standards and the use of fresh water for standards when samples are salt water based. It is thought that the uncertainty introduced by each of these processes is small compared to the pore water content and storage issues. Adoption of the

recommended procedures may reduce the potential for typical bias and total uncertainty in measurements to less than 5% and 15% respectively.

Table 8: Summary of uncertainty sources and magnitudes in order of analytical and data processing steps. B – bias, CV – coefficient of variation, U – total uncertainty. Typical Central Tendency (High) values are estimated from the data summarized in the relevant parts of this document. The values are meant to be illustrative of typical values and to help identify the most serious sources of uncertainty.

Analysis Process/Step	**Procedure and Uncertainty**	**Recommended Procedure and Uncertainty**
Pore Water Extraction	**Pore Water Not Extracted** B > 50% (0-100%); CV (unknown) U ~ 50% (0–100+%)	**Extract Pore Water** B ~0 ; CV (unknown) U ~ 0
Storage of Standards	**Titrated and Stored > 20 h** B > 20% (60%); CV ~5% (>50%) U ~20% (80%)	**Titrated Stock, Good Dilution of Standards and Storage 0h** B 1% (10%); CV 4% (11%) U 5% (14%)
Storage of Calibrated Electrodes	**Titrated and Stored < 4h** B -4% (~7%); CV ~4% (~7%) U ~5% (~10%) **Titrated and Stored > 4 h** B -15% (-22%); CV ~8% (16%) U 15% (25%)	**Titrated and Stored ~0h** B 1% (10%); CV 4% (11%) U 5% (14%)
Millivolt Reading	Meter (precision uncertain) Readings vary by ± 1 - <10 mV Readings taken within 2 min B >10% (>30%); CV 10% (50%) U ~20% (>80%)	**Meter (precision ~0.2 mV)** **Readings vary by ± 0.1 - 1 mV** **Readings taken within 2 min** B <1% (>1%); CV ~1% (8%) U ~2% (>9%)
Stock Solution	**No Titration** Depends on purity of chemical U ~20%	**Stock Solution Titrated** U ~0.1% (no info)
Generation of Dilution Series	**Poor Laboratory Procedure** U likely > 1%	**Good Laboratory Procedure** U < 1 % (no info)
Calibration Relationship	**Segmentation Used**	**Linear regression** (~5 pts)

Analysis Process/Step	Procedure and Uncertainty	Recommended Procedure and Uncertainty
	Sequential 2 points used Uncertainty estimates not available U >1%	Uncertainty generated from inverse regression

ENVIRONMENTAL CONTEXT

The variability in ISE derived sulfide measurements at the spatial scale of farms and reference locations is large and variable with a CV central tendency of greater than 60% and a range from less than 20% to greater than 100%. In contrast the variability in baseline or ideal ISE measurement precision is typically of order 10% i.e., considerably less than the coefficient of variation of in situ measurements.

There are many reasons for the increase in variability with distance between sub-samples and samples. Some of these include natural heterogeneity in sediment water content, patchiness in organic matter deposition and burial, resuspension that removes material in a non-uniform way from areas of high deposition immediately under net-pens and non-homogeneous redeposition, natural variation in amounts of FeS and other metal sulfides as well as variations associated with sediment sample collection (Hargrave pers. comm). Variability between analytical procedures used on samples from the same sampling program may also be a contributor to the spatial coefficients of variability; for example samples may not have been analyzed with the same time window and hence storage effects may have added to the spatial variability. Only rigorous standardized procedures that minimize analytical contributions to bias, enhance precision and hence enhance the total uncertainty can help ensure the robust detection of the environmental patterns.

IMPLICATIONS TO REGULATORY DECISION MAKING

The variability and bias in ISE sulfide measurements has implications for the consistency, and hence credibility, of regulatory results as well as how regulatory summary statistics should be calculated. This point has been made several times by various authors in the context of both finfish (Chang et al., 2011) and shellfish (Cranford et al., 2011) monitoring programs.

As an example, a comparison of empirical determinations of site mean sulfide concentrations was made using the data in Chang et al. (2011). The Chang et al. (2011) comparison was between mean ISE based sulfide concentrations determined by private industry as part of the New Brunswick Aquaculture Environmental Regulatory Program (NBDELG, 2018) and the concentrations determined by researchers; both means were based on samples collected from similar EMP locations but at different times and ISE sulfides were determined by different analysts and analytical equipment. The EMP and Chang et al. measurements both followed analytical procedures consistent with the method outlined in Wildish et al. (1999). The data indicate there is little relationship between repeated estimates of the site means and that the differences are sometimes large enough to correspond to different site classifications (Figure 13).

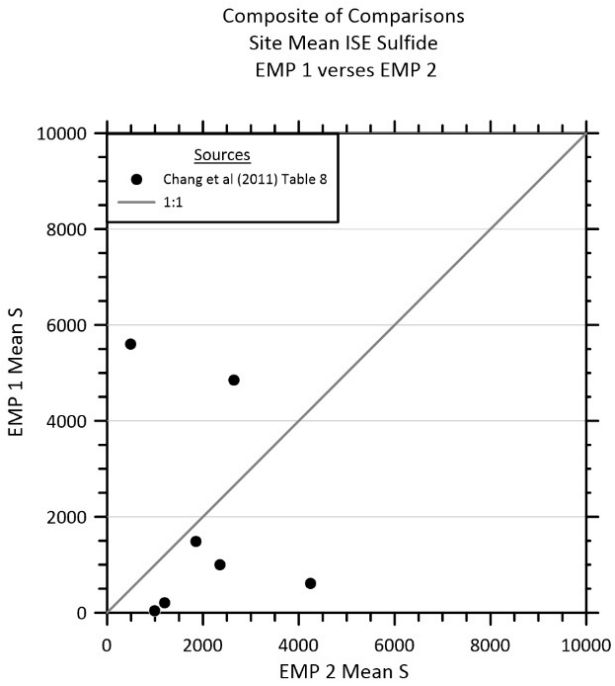

Figure 13: Composite scattergram showing Environmental Monitoring Program (EMP 1) site mean ISE based sulfide concentration verses a second determination of the site mean sulfide (EMP 2) based on an independent sampling and analyses of sulfide at stations similar to the EMP 1 stations.

The magnitude of the signal to be detected i.e., the difference between acceptable and unacceptable sulfide concentrations, is of order 1,000 μM (Appendix 3). The relative variation associated with rigorous ISE sulfide measurements is about 10% which corresponds to 10 μM for a 100 μM measurement, 100 μM for a 1,000 μM measurement, and 1,000 μM for a 10,000 μM. Since impacted sediments are considered to be those with ISE sulfide concentrations greater than 1,500 μM, the accuracy of the measurements is only slightly less than the magnitude of the signal to be detected.

The ability to statistically and robustly determine whether or not the sulfide conditions at a farm site exceed specific thresholds or differ from reference site conditions needs to be thoroughly explored. This ability is likely to depend on the site specific spatial, and perhaps temporal, variations in ISE and the efficiency of the sample design when a rigorous approach to ISE measurements, an approach that mitigates the factors that can cause significant bias and reduced precision in measured values, is implemented.

The evidence indicates that the variance in ISE sulfide measurements increases with the magnitude of the sulfide concentration. The calculation of arithmetic means or averages gives equal weight to each measurement, i.e., it assumes the uncertainty associated with each measurement is the same. However, in the case of ISE measurements the uncertainty is not constant. In this situation mean sulfide concentrations should be calculated as weighted means where the weights are defined by the variance associated with each measurement (Taylor, 1997). However, in existing regulatory sampling these variances are seldom known because multiple or replicate samples are not taken at each of the sample locations included in the averaging. An alternative approach is to log10 transform the estimated or measured sulfide values and take the average of these log transformed values. The log transformation normalizes the distribution of the sulfide values but appears not to fully stabilize the variability, i.e., observed variance in log10 transformed sulfide values has a slight negative relationship to

the mean of the transformed values (Figure 14). An alternative approach could be to use non-parametric statistical indicators such as a median. However, the regulatory thresholds would need to be reexamined and defined so thresholds are medians rather than means.

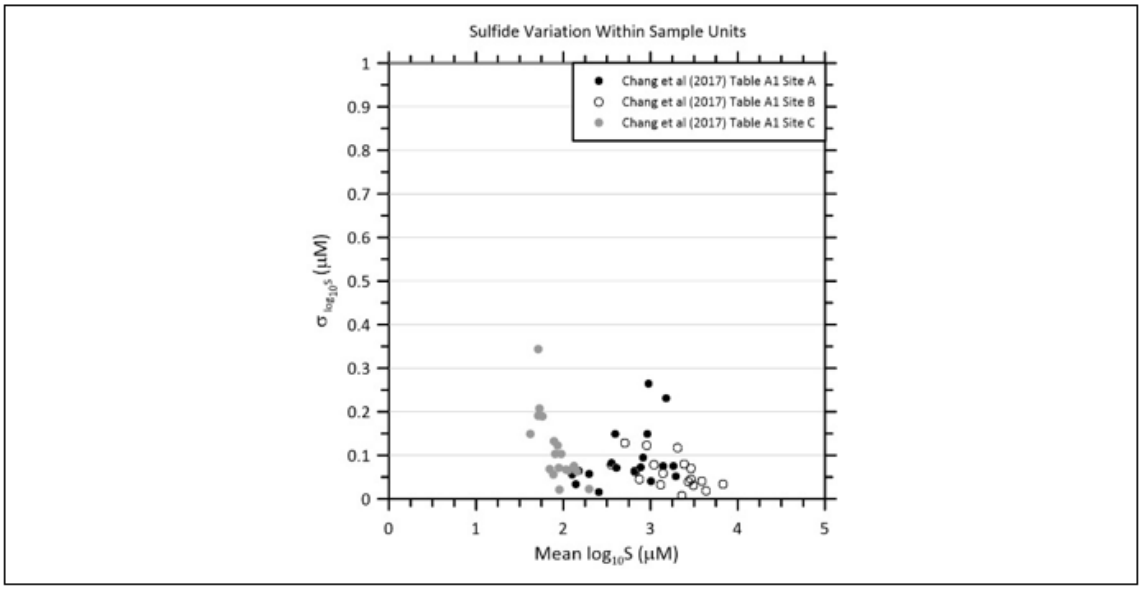

Figure 14: Example of the relationship between the standard deviation and mean of log10 transformed sulfide measurements. The data are from Table A1 of Chang et al. (2017).

In addition to viewing sulfide variability in the context of existing regulatory frameworks, the variability could be explored as an indicator of habitat heterogeneity and as an indicator of whether the aquaculture activity has changed this or as an indicator of the sampling effort needed to achieve the levels of uncertainty desired by regulators (suggested by B. Hargrave pers. comm.).

SECTION 6: SUMMARY AND CONCLUSIONS

Some specific summary bullets include:

1. All measurements have error, i.e., uncertainty associated with them, and ISE based sediment sulfide measurements are no different.
2. Measurement uncertainty can be considered as consisting of systematic (bias) and non-systematic (random) error. The random error is indicated by standard deviation and coefficient of variation indices.
3. Quantification of both systematic and non-systematic errors is essential if measurements are to have meaning and are to be interpreted with credibility.
4. The systematic error i.e., bias, associated with ISE sulfide measurements is variable, i.e., not consistent between experiments, but it is generally increased by storage of electrodes after calibration, storage of standard solutions, and storage of sediment samples.
5. The smallest bias and highest precision occurs when analyses were conducted very soon after sample preparation or collection, i.e., with almost zero storage time.
6. The bias when electrodes were not stored, i.e., used soon after calibration, was less than 5%. The bias increased with storage time so that within 4 h the bias was routinely greater

than but within ±5% and after 12 hours the bias was routinely more than 10%.The bias associated with standards and samples that are not stored is typically less than ±10%.

7. The bias associated with the storing of standards and samples is typically negative and can regularly exceed 10% when storage times exceed 4 h. At storage times longer than 4 h the bias magnitude and variability increase to more than 20% at times. Within these ranges the bias cannot be considered to be constant. It seems to vary with analyst, equipment, and sediment.
8. The precision of electrodes that are not stored is usually 5% or less. The bias for electrodes stored for 4 h or less after calibration is typically near 5% and biases associated with storage times greater than 12 h are consistently between 5 and 10%.
9. The precision (CV) associated with standards and samples than are not stored is centered around 5% and is typically less than 10%.
10. The precision associated with stored standards declines with storage time. After 24 hours the coefficient of variation can exceed 10% and after 48 to 72 h can exceed 20%. The largest CV was 45% after 72h of storage.
11. The precision associated with stored samples is variable but appears not to change with storage time.
12. Non-systematic error is random and only predictable in a probabilistic or statistical sense.
13. The frequency distribution of untransformed ISE sulfide concentrations is right (equal to positively) skewed, i.e., the data distribution tail extends to the right, toward the larger values. The mode is to the left.
14. Logarithmic transformation of ISE measurements results in less skewed error distributions. This is expected since ISE measurements are based on an inverse logarithmic relationship between mV and sulfide concentration. The frequency distribution of $\log_{10}$ transformed ISE sulfide concentrations approximates a normal distribution.
15. Indices of variability in ISE sulfide concentrations are not independent of the mean sulfide concentration. Variability in unhomogenized sediment sulfide concentration from field samples increases with the mean sulfide concentration. Logarithmic transformation of the ISE measurements reduces this dependence.
16. The largest source of non-systematic error in ISE sulfide measurements is associated with environmental inhomogeneity. Non-systematic error associated with ISE methodology technique is typically small compared to the environmentally induced variability.
17. The least biased and highest precision ISE measurements appear to be obtained when electrodes are calibrated immediately before use with freshly made standard solutions and samples are not stored.
18. Until more accurate measurement technologies are available, strict adherence to a well-defined and robust ISE methodology is essential to minimize measurement uncertainty. Methods that have a linear relationship between sulfide and the actual value that is measured should be preferred over those with non-linear relationships as in the case of ISE sulfide measurements unless the measurements and their manipulations are retained in the $\log_{10}$S form.

IMPLICATIONS TO DATA INTERPRETATION AND STATISTICAL INFERENCES

1. ISE measurements are subject to systematic (bias) and non-systematic uncertainties that data interpretation and decision making should properly take into consideration.

2. Parametric statistical comparisons between data from test, i.e., farm sites, and control or reference areas and between data from test areas and specified thresholds should be based on logarithmic transformations of the ISE sulfide measurements. This accommodates the realities that sulfide concentrations cannot be less than zero and that the ISE method is based on a logarithmic calibration relationship.
3. Non-parametric statistical comparisons between test and control areas or specified thresholds are not sensitive to the normality assumption and are preferred for low sample number datasets. The non-parametric statistics considered here, Wilcoxon Signed-Ranks test, however assumes symmetry in the data distribution.
4. At least five samples should be taken from test areas in order to statistically detect exceedances of thresholds at the alpha 0.05 level, i.e., to attribute the test results occurring less than 5% of the time if the null hypothesis of no exceedance is true. More samples will provide a more robust detection of differences. This only applies if the samples are not biased.
5. Sulfide values do not need to be transformed if non-parametric tests are to be used in hypothesis testing, i.e., results will be the same if sulfide values are transformed or not because tests are based on value rankings which are unaffected by value transformation.
6. The measurement precision attributable to technique represents a small proportion of field measurement variability; most of the variability in field measurements is related to systematic and non-systematic heterogeneities in the environment. The bias that may be associated with measurements that have been made on stored samples or have been made with stored electrodes or standard concentrations can result in a shift in measurement values that can affect interpretations of site classification.
7. Sampling designs need to consider the realities of the magnitude of localized environmental error.
8. Data interpretation and decision-making needs to properly account for measurement uncertainties and sampling designs.

SECTION 7: RECOMMENDATIONS

1. To eliminate the potential for bias and increases in variability do not store standards and samples.
2. Conduct a review of the updated and detailed statement of standard operating procedures (SOP) to determine if they can be practically implemented (technically feasible, cost-effective, etc.) by those conducting regulatory, auditing, and supportive research analyses.
    - If adjustment of ISE sulfide concentrations for water content is included in a revised SOP, consideration may need to be given to re-establishing the sulfide thresholds since these were established using unadjusted sulfide concentrations.
    - If it is not feasible to eliminate storage of standards and samples, then interpretations of monitoring results will need to consider that a variable bias of between 10% and 50% or more is likely.
3. Examine the strengths and weaknesses of ISE based measurements in the context of potential alternative measurement methods and regulatory thresholds to determine if more suitable alternatives exist. This should include considerations of pore water extraction and analytical methodologies. Pore water extraction has the potential to minimize the uncertainties introduced by sediment matrix and water content effects.

4. Regulatory monitoring programs and sampling designs should be reviewed to determine if they appropriately and adequately account for the measurement uncertainty associated with the ISE methodology and associated sampling designs and decision criteria. For example, should practical thresholds be established that are one or two standard deviations or standard errors below the thresholds of most concern so the probability of false threshold exceedances are reduced.

SECTION 8: ACKNOWLEDGEMENTS

The authors benefitted from several conversations and literature insights provided by B.D. Chang at various times throughout the preparation of this document. The authors also thank the multiple reviewers of this document for their comments and insights as these significantly improved the content of the document. The suggestion to include a summary table of bias and precision effects has especially improved the manuscript.

SECTION 9: REFERENCES CITED

Berendsen, H.J. 2011. A Student's Guide to Data and Error Analysis. Cambridge University Press.

Bugden, J.B., Hargrave, B.T., Strain, P.M., and Stewart, A.R. 2001. Spatial Patterns of Some Physical and Chemical Variables in Passamaquoddy Bay and Letang Inlet, Southwstern Bay of Fundy, September 1999. Can. Tech. Rep. Fish. Aquat. Sci. 2356: iv + 96 p.

Chang, B.D., Page, F.H., Losier, R.J., McCurdy, E.P., and MacKeigan, K.G. 2011. Characterization of the spatial pattern of benthic sulfide concentrations at six salmon farms in southwestern New Brunswick, Bay of Fundy. Can. Tech. Rep. Fish. Aquat. Sci. 2915: iv + 24 p.

Chang, B.D., Balch, T., Bennett, A.T., Buchan, C.M., Page, F.H., Jacobi, C.A., Bagnall, A.G. 2013. Spatially-intensive sediment surveys at two salmon farms in New Brunswick and Nova Scotia in September-October 2011. Can. Tech. Rep. Fish. Aquat. Sci. 3035: v + 32 p.

Chang, B.D., Bennett, A.T., Lyons, T.A., Parker, E.V., and Page, F.H. 2013b. Analysis of sediment sulfide concentration data from Environmental Management Program Tier 2 monitoring at salmon farms in southwestern NEw Brunswick. Can. Tech. Rep. Fish. Aquat. Sci. 3033: v + 49 p.

Chang, B.D., Lewis-McCrea, L.M., Wong, D.K., MacKeigan, K.G., Page, F.H., Cameron, S.L., Smith, A.E. 2014. InterLaboratory research on factors affecting the calibration of electrodes used for sediment sulfide analyses at fish farms. Can. Tech. Rep. Fish. Aquat. Sci. 3090: v + 66 p.

Chang, B.D., Cooper, J.A., Page, F.H., and Losier, R.J. 2017. Benthic macrofaunal biodiversity in relation to sediment sulfide concentration under salmon farms in southwestern New Brunswick, Bay of Fundy. Can. Tech. Rep. Fish. Aquat. Sci. 3202: v + 71 p.

CoastalSmith. 2002. Standard Operating Practices for the Monitoring Program of the Environmental Management Guidelines for the Marine Finfish Cage Aquaculture Industry in New Brunswick. CoastalSmith, Inc., 26 p.

Cranford, P.J., Balch, T., Buchan, C., Crocker, J., Filgueira, R., Grant, J., Stuart, R. 2011. Mussel aquaculture regulatory effectiveness monitoring: Validation of the St. Ann's Harbour environmental assessment and monitoring program. Can. Tech. Rep. Fish. Aquat. Sci. 2905: vi + 25 p.

Cranford, P., Brager, L., Elvines, D., Wong, D., and Law, B. 2020. A revised classification system describing the ecological quality status of organically enriched marine sediments based on total dissolved sulfides. Mar. Poll. Bull. 154.

DFO. 2018. AAR Monitoring Standard 2018. Fisheries and Oceans Canada. Retrieved from http://www.dfo-mpo.gc.ca/aquaculture/management-gestion/aar-raa-eng.htm

Fischer Scientific. 2014. Eletrochemistry Handbook. Thermo Fischer Scientifc Inc.

Hall, B.D., and White, D.R. 2020. An Inroduction to Measurement Uncertainty. Measurement Standards Laboratory of New Zealand.

Hargrave, B.T., Doucette, L.I., and Milligan, T.G. 1993a. Geochemical Characteristics and Benthic Macrofauna Biomass in Intertidal and Subtidal Sediments of Annapolis Basin, Nova Scotia, 1993. Can. Data Rep. Fish. Aquat. Sci. 915: iv + 88 p.

Hargrave, B.T., Duplisea, D.E., Pfeiffer, E., and Wildish, D.J. 1993b. Seasonal changes in benthic fluxes of dissolved oxygen and ammonium associated with marine cultured Atlantic salmon. Mar. Ecol. Prog. Ser. 96: 249-257.

Hargrave, B.T., Phillips, G.A., Doucette, L.I., White, M.J., Milligan, T.G., Wildish, D.J., and Cranston, R.E. 1995. Biogeochemical observations to assess benthic impacts of organic enrichment from marine aquaculture in the Western Isles region of the Bay of Fundy, 1994. Can. Tech. Rep. Fish. Aquat. Sci. 2062: v + 159 p.

Hargrave, B.T., Phillips, G.A., Doucette, L.I., White, M.J., Milligan, T.G., Wildish, D.J., and Cranston, R.E. 1997. Assessing Benthic Impacts of Organic Enrichment From Marine Aquaculture. Water, Air and Soil Pollution. 99: 641-650.

Hargrave, B.T., Doucette, L.I., Phillips, G.A., Milligan, T.G., and Wildish, D.J. 1998. Biogeochemical observations to assess benthic impacts of organic enrichment from marine aquaculture in the Western Isles region of the Bay of Fundy, 1995. Can. Data Rep. Fish. Aquat. Sci. 1031: iv + 50p.

Hargrave, B.T., Doucette, L.I., Cranford, P.J., Law, B.A., and Milligan, T.G. 2008a. Influence of mussel aquaculture on sediment organic enrichment in a nutrient-rich coastal embayment. Mar Ecol Prog Ser. 365: 137–149.

Hargrave, B.T., Holmer, M., and Newcombe, C.P. 2008b. Towards a classification of organic enrichment in marine sediments based on biogeochemical indicators. Mar. Poll. Bull. 810-824.

Maley, C.H. 2012. Project Management Concepts, Methods, and Techniques. CRC Press, Taylor Francis Group.

NBDELG. 2004. Standard Operating Practices for the Environmental Monitoring Program of the Environmental Management Guidelines for the Marine FInfish Cage Aquaculture Industry in New Brunswick. New Brunswick Department of Environment and Local Government (NBDELG), 25 p.

NBDELG. 2018. Standard Operating Practices for the Environmental Monitoring of the Marine Finfish Aquaculture Industry in New Brunswick. New Brunswick Department of Environment and Local Government, 38 p.

NICO. 2000. Frequently asked questions about ion-selective electrode measurements. Retrieved 11 26, 2017, from www.nico2000.net/datasheets/faq.html#2

NSFA. 2011. Environmental Monitoring Program Framework For Marine Aquaculture In Nova Scotia. Nova Scotia Fisheries and Aquaculture (NSDA).

Page, F.H., Chang, B.D., Losier, R.J., McCurdy, E.P., Reid, J.C., Hanke, A.R., and Cooper, A.J. 2010. Temporal changes in sediment sulphide levels under salmon farms in southwestern New Brunswick, Bay of Fundy, during the autumn environmental monitoring period. Presentation to the AAC.

Page, F.H., Chang, B.D., Losier, R.J., McCurdy, E.P., Reid, J.C., and Hanke, A.R. 2011. Temporal variations in sediment sulfide levels under marine salmon farms in southwestern New Brunswick, Bay of Fundy, during the annual environmental monitoring period. p. 64-66 in Aquaculture Canada 2010 and Cold Harvest 2010 – Proceedings of the Contributed Papers, AAC Special Publication No. 17.

Qian, S.S. 2010. Environmental and Ecological Statistics with R. Chapman and Hal/CRC, Taylor and Francis Group.

Rowe, S.F. 2015. Project Management for Small Projects Second Edition. Management Concepts Press.

Rundle, C.C. 2011. A Beginners Guide to Ion-Selective Electrode Measurements. London, UK.: NICO2000 Ltd. Retrieved 11 26, 2017, from www.nico2000.net

Rundle, C.C. 2011. A Beginners Guide to Ion-Selective Electrode Measurements. Nico2000 Ltd, London, UK.

Schaaning, M.T., and Hansen, P.K. 2005. The suitability of electrode measurements for assessment of benthic organic impact and their use in a management system for marine fish farms. In B. T. Hargrave, Environmental Effects of Marine Finfish Aquaculture. Handbook of Env. Chem. Vol. 5 Part M (2005) (pp. 381-408). Springer-Verlag Berlin Heidelberg.

Sheskin, D.J. 2007. Handbook of Parametric and Nonparametric Statistical Procedures, Fourth Edition. Chapman and Hall/CRC.

Taylor, J.R. 1997. An Introduction to Error Analysis The Study of Uncertainties in Physical Measurements Second Edition. Sausalito, California: University Science Books.

TFS. 2007. User Guide Silver/Sulfide Ion Selective Electrode. Thermo Fisher Scientific Inc.

van Belle, G. 2002. Statistical Rules of Thumb. John Wiley and Sons, Inc. Wiley Series in Probability and Statistics.

Wildish, D.M., Akagi, H.M., Hamilton, N., and Hargrave, B.T. 1999. A recommended Method for Monitoring Sediments to detect Organic Enrichment from Mariculture in the Bay of Fundy. Can. Tech. Rep. Fish. Aquat. Sci. 2286: iii + 31 p.

Wildish, D.J., B.T. Hargrave and G. Pohle. 2001. Cost-effective monitoring of organic enrichment resulting from salmon aquaculture. ICES J. Mar. Sci. 58: 469-476

Wildish, D.J., Akagi, H.M., and Martin, A. 2002. Seasonal Patterns of Biological and Physical Variables in Sediments of Lime Kiln Bay during 2000-2001. Can. Tech. Rep. Fish. Aquat. Sci. 2447: iii + 46 p.

Wildish, D.W., Akagi, H.M., Hargrave, B.T., and Strain, P.M. 2004a. Inter-laboratory Calibration of Redox Potenital and Total Sulfide Measurements in Interfacial Marine Sediments and the Implications for Organic Enrichment Assessment. Can. Tech. Rep. Fish. Aquat. Sci. 2546: iii + 25 p.

Wildish, D.J., M. Dowd, T.F. Sutherland and C.D. Levings. 2004b. A scientific review of the potential environmental effects of aquaculture in aquatic ecosystems. Volume III. Near-field organic enrichment from marine finfish aquaculture. Can. Tech. Rept. Fish. Aquat. Sci. 2450: ix + 117

Wildish, D.J., Akagi, H.M., and Hamilton, N. 2005. Interfacial Geochemistry and Macrofauna at a New Salmon Farm in Passamaquoddy Bay, Bay of Fundy. Can. Tech. Rep. Fish. Aquat. Sci. 2574: iii + 40 p.

Wong, D., Page, F.H. 2025a. Ion Selective Electrode (ISE) Method Overview and Variations in Implementation Descriptions for the Determination of Sulfide in Marine Sediment Samples. DFO Can. Sci. Advis. Sec. Res. Doc. 2023/015. iv + 47 p.

Wong, D., and Page, F. 2025b. Storage Stability: Effect of Storage Time, Temperature and Preservation Method on Total Free Sulfide Measurements in Marine Benthic Sediment. DFO Can. Sci. Advis. Sec. Res. Doc. 2025/016. iv + 43 p.

Zar, J.H. 1999. Biostatistical Analysis Fourth Edition. Prentice Hall.

APPENDIX 1: DERIVATION OF SEDIMENT WATER CONTENT ADJUSTMENT FOR ISE SULFIDE

The calibration equation associated with the ISE total free sulfide method is a relationship between the millivolts (mV) associated with a solution of sulfide and buffer that has known concentrations (S_{NC}) i.e, the nominal concentrations of the standard solutions. The nominal concentrations are prepared as a dilution series from a stock solution whose concentration is established by titration (Wong and Page, this CSAS). Volumes of the diluted concentrations are then diluted when they are mixed with the L-ascorbic acid and SAOB buffer (referred to as SAOB hereafter). The concentration of sulfide in the solution that the ISE electrode is immersed in is given by

$$S_{cal\ st} = \frac{S_{NC}\ x\ V_{NC}}{V_{NC} + V_{SAOB}}$$

In the ISE method described by Wildish et al. (2004a) and Wong and Page (this CSAS), the volume (5 ml) of each sulfide calibration standard (V_{NC}) is specified to be equal to the volume (5 ml) of the SAOB (V_{SAOB}), i.e., the ratio (R) of the volume of sulfide standard to the volume of SAOB is 1:1 or $R = V_{NC}/V_{SAOB} = 1$. The millivolts associated with the calibration curve are therefore associated with a standard concentration that is half the nominal concentration i.e.,

$$S_{cal\ st} = \frac{S_{NC}\ x\ V_{NC}}{V_{NC} + V_{SAOB}} = \frac{S_{NC}\ x\ 5ml}{5ml + 5ml} = \frac{S_{NC}}{2} = 0.5S_{NC}$$

The ISE sulfide calibration equation is therefore

$$mV = \beta_0 + \beta_1 log_{10}(0.5S_{NC}) = \beta_0 + \beta_1 log_{10}(0.5) + \beta_1 log_{10}(S_{NC})$$

By setting $\beta_0 + \beta_1 \log(0.5)$ to β_0', the calibration equation, the equation that is used, can be rewritten as

$$mV = \beta_0' + \beta_1 log_{10}(S_{NC})$$

The direct use of the equation is appropriate when the ISE electrode is inserted into solution that consists of equal volumes of solution and SAOB, since the millivolts recorded are converted into the sulfide concentration through the use of the calibration equation. The values of the relationship coefficients (β_0' and β_1) are derived from the calibration procedure millivolt and standard nominal sulfide solution concentration data through the segmented approach or a regression approach. This is the situation described in the analyses reported in Chang et al. (2014).

The direct use of the equation is not appropriate when the ISE electrode is inserted into a solution that consists of unequal volumes of solution and SAOB. The application of the ISE sulfide method for determining sediment pore water sulfide concentration is a case of unequal volumes of solution and SAOB. In this case the ISE electrode is immersed into a solution that is composed of a volume of sediment (V_{sed}) mixed with a volume of SAOB. The volume of SAOB is the same as that used in the deriving the calibration relationship but the volume of the test solution, the sediment (V_{sed}) is not all water; it consists of inorganic and organic substances plus pore water. The volume of water (V_{pw}), termed the pore water, is given by $V_{pw} = V_{sed}\ x\ WC/100$ when WC is the water content in units of percent. When the millivolt reading obtained from the immersion in the sediment plus SAOB solution is the result of the sulfide ion activity in the pore

water modified by any sediment – pore water matrix effects that may be present. The typical approach is to assume the matrix effects are of no consequence and that the millivolt reading is representative of the sulfide concentration in the sediment volume i.e., the sulfide ions in the pore water diluted throughout the sediment volume. To correct for this effect the ISE based sulfide concentration (S_{ISE}) needs to be adjusted to account for the fact the sulfide ions are actually only in a fraction of the sediment volume. Hargrave et al. (2010) have suggested the sulfide concentration in the pore water (S_{pw}) is given by

$$S_{pw} = \frac{Q_S}{V_{pw}} = \frac{S_{ISE}\ x\ V_{sed}}{V_{sed} x\ WC/100} = S_{ISE}\ x\ 100/WC$$

This relationship is based on the logic that the quantity of sulfide (Q_S) in the sediment volume, $Q_S = S_{ISE}\ x\ V_{sed}$, is estimated by the calibration derived estimate of the sulfide concentration, S_{ISE}, times the volume of the sediment (V_{sed}). By assuming this sulfide is actually within the volume of the pore water (V_{pw}), the concentration of sulfide within the pore water (S_{pw}) is given by the above relationship.

APPENDIX 2: PRINCIPLES OF MEASUREMENT UNCERTAINTY

SYSTEMATIC ERROR OR BIAS

The systematic error ($S_{\varepsilon b}$) is constant under defined circumstances and has, at least in theory, the potential to be known. Systematic error is consistent and is either an over or under estimation of the true value. The magnitude of the bias (*B*) is indicated by the absolute (Eq. A2-1) or relative (Eq. A2-2) difference between the average measured value ($\bar{S}_m$) and the true value (S_T). As indicated above, the error associated with every individual measurement includes bias and random uncertainty. Bias cannot be determined from single measurements; it can only be estimated from multiple measurements made on a single or homogeneous sample in a consistent manner. Unfortunately, in situ sediments are generally not homogeneous.

Bias in ISE measurements are difficult to determine because the true value is seldom known. Estimates of measurement bias can therefore only be derived from samples that have been prepared to have a known concentration of sulfide.

$$B_R = (\bar{S}_M - S_T)/S_T \quad \text{Eq. A2-2}$$

$$B_A = \bar{S}_M - S_T \quad \text{Eq. A2-1}$$

NON-SYSTEMATIC ERROR OR RANDOM UNCERTAINTY

The random error ($S_{\varepsilon r}$) associated with single measurements cannot be predicted, only the statistical properties of the random error can be known. The random error in sulfide measurements can be indicated by several statistics including the range (R_S), standard deviation (σ_S), standard error and coefficient of variation (CV_S). The statistics are calculated from multiple measurements taken in consistent ways. The range (Eq. A2-3) is calculated as the maximum value minus the minimum values, the unbiased standard deviation (Eq. A2-4) is calculated as the square root of the sum of the squared differences between individual measured values and the mean divided by the number of samples (n) minus one, the coefficient of variation (Eq. A2-5) is calculated as the standard deviation divided by the mean and the standard error (Eq. A2-6) is calculated as the standard deviation divided by the square root of the number of samples.

$$R_S = \max(S) - \min(S) \quad \text{Eq. A2-3}$$

$$\sigma_S = \left(\sum_{i=1}^{n} (S_i - \bar{S})^2/(n-1)\right)^{0.5} \quad \text{Eq. A2-4}$$

$$CV_S = \sigma_S/\bar{S} \quad \text{Eq. A2-5}$$

$$SE_{\bar{S}} = \sigma_S/\sqrt{n} \quad \text{Eq. A2-6}$$

TOTAL UNCERTAINTY

Total uncertainty (*U*) is the combination of systematic error or bias (*B*) and non-systematic uncertainty (σ) or random error (Taylor, 1997) and is indicated by estimates of the absolute uncertainty (U_A, Eq. A2-7) or relative uncertainty (U_R, Eq. A2-8).

$$U_A = \sqrt{(B_A)^2 + (\sigma_S)^2} \quad \text{Eq. A2-7}$$

$$U_R = \sqrt{(B_R)^2 + (CV)^2} \quad \text{Eq. A2-8}$$

Uncertainty is often considered using the terms accuracy and precision. Accuracy embodies the concept of how close a measurement is to its true value and includes the concept of precision. Accuracy is indicated by the total uncertainty (U_A *or* U_B). Precision embodies the concept of how similar repeated measurements of the same entity are or how close together the measurements are (Zar, 1999). Precision is often considered as the scatter in repeated measurements and the scatter is usually considered to be random and unpredictable. Precision is indicated by the measurement variability (R_S, σ_s, CV_S *or* $SE_{\bar{S}}$).

The International Organization for Standardization (ISO) considers an accurate measurement to be one in which the central tendency is very close and ideally identical, to the true value and the scatter or variation is very small and ideally zero. In other words, both the bias and random error are small or non-existent and hence the total uncertainty is small or non-existent.

In the case of environmental measurements such as ISE sulfide concentrations both the systematic and non-systematic error may vary with space and time. The measurements can potentially be adjusted to account for systematic error (bias) when the error has been quantified for the particular measurement conditions because the bias is constant and predictable under these conditions. Individual measurements cannot be adjusted to account for random error because it is unpredictable. However, when the properties of the random error are quantified, sample design can help account for the uncertainties and lead to efficient and statistically proper assumptions when making probabilistic inferences from the measurements.

In the following sections we explore the magnitudes of the systematic (bias) and non-systematic (random) uncertainties associated with the ISE method.

UNCERTAINTY ASSOCIATED WITH AVERAGING ISE MEASUREMENTS

As stated above ISE measurements (S_M), like all measurements, consist of a true (S_T), but usually unknown value, plus an uncertainty ($U_{\epsilon T}$, Eq. A2-9).

$$S_M = S_T + U_{\epsilon T} \quad \text{Eq. A2-9}$$

The total uncertainty in ISE sulfide measurements ($U_{\epsilon T}$) consists of a systematic error ($U_{\varepsilon b}$) or biases (*B*) plus non-systematic or random error ($U_{\epsilon r}$, Eq. A2-10).

$$U_{\varepsilon T} = U_{\varepsilon b} + U_{\varepsilon r} \quad \text{Eq. A2-10}$$

When multiple sulfide measurements are averaged, the mean sulfide ($\bar{S}_M$) has a total error ($\bar{\varepsilon}_T$) associated with it (Eq. A2-11) and when the error is constant, $\bar{\varepsilon}_T = \varepsilon_T$.

$$\begin{aligned} \bar{S}_M &= \frac{\sum_{i=1}^{n}\left(S_{M,i} \pm \varepsilon_T\right)}{n} \\ &= \frac{\sum_{i=1}^{n}\left(S_{M,i}\right) \pm \sum_{i=1}^{n}(\varepsilon_T)}{n} \\ &= \bar{S}_M \pm \bar{\varepsilon}_T \end{aligned} \quad \text{Eq. A2-11}$$

When the error is not constant the averaging needs to adjust for this. One way to do so is to take a weighted mean (see main text).

APPENDIX 3: SUMMARY OF CANADIAN SULFIDE THRESHOLDS

Table A3- 1: Summary of sediment pore water ISE sulfide thresholds with some additional information. The log_{10} sulfide and millivolt columns are not in the source document. The millivolt levels are provided as an example since the exact millivolt levels will vary with each calibration. The example millivolts were calculated assuming the example calibration relationship $mV = -798.0 -29.5log_{10}(S)$ where the intercept and slope are rounded to the first decimal place.

Sediment Classification	Qualitative Description	Sulfide (μM)	Sulfide $Log_{10}(S)$	Millivolts (mV)
Background				
Oxic	Typical background or normal levels[2]	<300[2]	<2.48	< -871.0
Thresholds Indicating Impact Relative to Oxic				
Oxic A	Low effects	<750	<2.88	-871 to < -882.7
Oxic B	Low effects	750 to <1 500	2.88 to <3.18	-882.7 to < -891.6
Post Oxic	Transitory[2]	300 to 1 300[2]	-	-
Hypoxic A	May be causing adverse effects	1 500 to <3 000	3.18 to <3.48	-891.6 to < -900.5
Hypoxic B	Likely causing adverse effects	3 000 to <4 500	3.48 to <3.65	-900.5 to < -905.7
Hypoxic C	Causing adverse effects	4 500 to <6 000	3.65 to 3.78	-905.7 to < -909.4
Anoxic	Causing severe damage	>6 000	>3.78	> -909.4
Other Threshold Definitions				
Oxic	Normal[3]	<50[3]	-	-
Oxic	Transitory[3]	>50[3]	-	-
Post-Oxic[2]	Transitory[2]	300 to 1 300[2]	-	-
Hypoxic	Polluted[3]	>200[3]	-	-
Hypoxic[2] Sulfidic[2]	Polluted[2]	1 300 to 6 000[2]	-	-

Sediment Classification	Qualitative Description	Sulfide (μM)	Sulfide $Log_{10}(S)$	Millivolts (mV)
Anoxic	Grossly polluted[3]	>2 000[3]	>3.30	-895.3
Anoxic[2] Methanic[2]	Grossly polluted[2]	>6 000[2]	-	-
Differences between Thresholds				
anoxic - background	Background to Severe Damage	5 700	1.30	-38.4
Hypoxic B – Hypoxic A	Likely causing adverse effects to May be causing damage	1 500	0.30	-8.9
Hypoxic A – Hypoxic C	May be causing damage to Causing adverse effects	3 000	0.48	-14.1
Hypoxic A – Anoxic	May be causing damage to Severe damage	>4 500	0.60	-17.8

[1] - It should be recognized that not all background environments are oxic with low sulfides; some areas have naturally high sulfides.
[2] - Hargrave et al. (2008b);
[3.] - Hargrave et al. (1997)

Provincial regulatory environmental impact classifications in Canada vary between provinces. The requirements relating to sediment sulfide are summarized in the following tables. The regulatory interpretation of environmental effects associated with sediment sulfide measurements has changed over the years in New Brunswick. The most recent regulatory categories recognize oxic, hypoxic and anoxic classifications and no longer recognize the subdivisions within these classifications (Table A-NB2).

Table A3-NB1. Finfish aquaculture site classifications in New Brunswick based on mean sediment sulfide concentration in annual monitoring of finfish farms in New Brunswick based on NBDELG (2012a) as reported in Chang et al. (2013).

Site Classification	Qualitative Description	Site Mean Sediment Sulfide
Oxic A	low effects	<750 μM
Oxic B	low effects	750 to 1499 μM
Hypoxic A	may be causing adverse effects	≥1 500 to 2 999 μM
Hypoxic B	likely causing adverse effects	≥3 000 to 4 999 μM

Site Classification	Qualitative Description	Site Mean Sediment Sulfide
Hypoxic C	causing adverse effects	≥4 999 to 5 999 μM
Anoxic	causing severe damage	>6 000 μM

Table A3-NB2: Finfish marine net pen aquaculture sediment sulfide site classification criteria in New Brunswick, Canada in effect as described on the NB Website on April 2, 2022. Source given below table.

Site Classification	Qualitative Description	Site Mean Sediment Sulfide	Regulatory Action
Oxic	good performing site	<1 500 μM	minimal remedial action required.
Hypoxic	sites are likely to cause adverse environmental affects	≥1 500 to 6 000 μM	required to undertake a number of remedial measures.
Anoxic	sites are causing severe damage to the marine habitat.	>6 000 μM	sites directed to work closely with regulatory agencies to remedy the situation.

Source: information as given on the New Brunswick Canada Marine Aquaculture site accessed on 2 April 2022.

Table A3-NS: Finfish aquaculture sediment sulfide concentration classifications in Nova Scotia. Information obtained from Figure 1 in (NSFA, 2011).

Site Classification	Mean Sediment Sulfide (total S^{2-})	Effects on marine sediments	Management Action Thresholds
Oxic A	< 1 500 μM	Low effects	Not Hypoxic or Anoxic
Oxic B	> 750 and < 1 500 μM	Low effects	
Hypoxic A	> 1 500 μM and <3 000 μM	Likely causing adverse effects near some cage structure	≥ 50% of stations with means > 1 500 μM
Hypoxic B	> 3 000 μM and < 6 000 μM	Likely causing adverse effects near some cages	≥ 50% of stations with means > 3 000 μM

Site Classification	Mean Sediment Sulfide (total S^{2-})	Effects on marine sediments	Management Action Thresholds
Anoxic	> 6 000 µM	Causing adverse effects to large portions of site	≥ 70% of stations

Table A3- 2: Summary of the number of sampling stations and sample numbers required for ISE sediment sampling in (Nova Scotia and New Brunswick) according to the Canadian Federal Aquaculture Activities Aquaculture Monitoring Standard (DFO, 2018). This is a reproduction of Table 1 in the monitoring standard.

Maximum number of fish within cage site array during production cycle	**Number of transects**	**Number of sampling stations (not including reference stations)**	**Number of samples (3 samples/station for soft bottom sites at 0 m from containment array)**
1-200,000	2	2	6
200,001-300,000	3	3	9
300,001-400,000	4	4	12
400,001-500,000	4	5	15
500,001-600,000	4	6	18
600,001-700,000	4	7	21
700,001-800,000	4	8	24
800,001-900,000	4	9	27
900,000-1,000,000	4	10	30

APPENDIX 4: SUMMARY OF BASIC STATISTICS

In this appendix some of the basic statistics used in the document are listed (Table A4-1) along with a brief explanation of what each represents and the formula(s) used to calculate their values. A simulated set of measurements is given in Table A4-2 along with the calculated values of the statistics.

Table A4-1: A summary of some of the basic statistics referred to in this document. The SD and CV statistics are most meaningful for Normal distributions and should be combined with other spread estimates for non-normal data such as ISE sulfide data.

Statistic	Comment	Formula
Bias or B	The bias in sample values. An estimate of the systematic difference between the true sulfide value and the mean of measured sample values.	$B = \bar{S} - S_T$
CV	The coefficient of variation is a relative measure of relative measurement variability. About ~68%, 95% and 99.7% (almost all) of individual measurements fall within ±1CV , ±2CV and ±3CV of the sample mean. Estimates of relative precision.	$CV = \frac{\sigma}{\bar{S}}$ $2CV = 2\sigma/\bar{S}$
i	The sample number.	i
Mean or $\bar{S}$	The arithmetic average of the measurements. When the samples are from a symmetrical distribution ~50% of the values are greater than the mean and ~50% are less than the mean.	$\bar{S} = \sum_{i=1}^{n} S_i$
Min	The smallest sample measurement value.	
Max	The largest sample measurement value.	
MSE	The mean squared error. This is the average of the squared differences between sample values and the true value or its best estimate. An estimate of accuracy.	$MSE = \frac{\sum_{i=1}^{n}(S_i - S_T)^2}{n}$ $MSE = B^2 + \sigma^2$
RMSE	The root mean squared error. An estimate of accuracy.	$RMSE = \sqrt{MSE}$
n	The number of sample measurements.	n
S	A sample sulfide value.	S_i
S_T	The true sample value. This is usually unknown.	S_T
$\bar{S}$	see Mean	

Statistic	Comment	Formula
SD or σ	The standard deviation of sample measurements. This indicates the absolute spread in values around the sample mean. An estimate of absolute precision. About 68%, 95% and 99.7% (almost all) of individual measurements fall within ±1σ , ±2σ and ±3σ of the sample mean. Estimates of relative precision.	$\sigma = \sqrt{\sigma^2}$
SE	The standard error of the estimated sample mean. This gives an indication of the spread of mean values if multiple sampling exercises were to be taken. When the sample values are from a Normal distribution the true mean should be within two SE of the sample mean about 95% of the time.	$SE = \sigma/\sqrt{n}$
Range or R	The spread in the sample values. An estimate of the absolute precision.	$\max(S_i) - \min(S_i)$
U	The total uncertainty. This is equivalent to the RMSE. It includes absolute bias and precision.	$U = MSE = B^2 + \sigma^2$
Var or σ^2	The variance of sample measurements. This indicates the spread in values around the sample mean and is an indication of precision. Division by *n-1* is the unbiased estimate of variance.	$\sigma^2 = \frac{\sum_{i=1}^{n}(S_i - \bar{S})^2}{n}$ $\sigma^2 = \frac{\sum_{i=1}^{n}(S_i - \bar{S})^2}{n-1}$

Table A4-2: A simulated data set with values of various statistics that have been calculated from the simulated data. MSE[1] and RMSE[1] have been calculated directly from the simulated values using the formulas indicated in Table A- 1. MSE[2] and RMSE[2] have been calculated using the MSE = Bias squared – Variance relationship. Simulation input parameters are shown in (a). Values, value statistics, bias and precision are shown in (b). Note that 3CV @ Range, MSE[1] = MSE[2], RMSE[1] = RMSE[2], and true value is outside range of sample estimated mean value.

(a)

Simulation Input Parameters				
True Value (mM)	**True Bias (mM)**	**True Bias (%)**	**True SD (µM)**	**True CV (%)**
2000	200	10.0	100	4.5

(b)

Values (mM)	Value Statistics		
	Statistic	**Value**	**Units**
2003	n	10	N/A
2046	mean	2118	mM
2035	**Bias (systematic error)**		
2182	bias	118	mM
2078	bias	5.9	%
2166	SE	22	mM
2143	**Precision (random error)**		
2228	min	2003	mM
2124	max	2228	mM
2176	range	225	mM
-	s^2	4937	mM^2
-	s	70.3	mM
-	2s	140.5	mM
-	3s	210.8	mM
-	CV	3.3	%
-	2CV	6.6	%
-	3CV	10.0	%
-	MSE^1	18863	mM^2
-	$RMSE^1$	137	mM
-	MSE^2	18863	mM^2
-	$RMSE^2$	137	mM
-	SE	22.2	mM
-	2SE	44.4	mM
-	3SE	66.7	mM
-	mean-3SE	2051.4	mM

Values (mM)	Value Statistics		
	Statistic	Value	Units
-	mean+3SE	2184.7	mM

APPENDIX 5: PROPAGATION OF MILLIVOLT UNCERTAINTY IN THE CALIBRATION RELATIONSHIP

The ISE calibration equation (Eq. A5- 1) relates millivolts to log transformed sulfide concentration. Estimates of unknown transformed sulfide concentration are derived from the inverse of the calibration equation (Eq. A5- 2).

The accuracy of millivolt measurements associated with the ISE method varies for several reasons including; the millivolt meter may have limited accuracy, the analyst using the meter may affect the reading by subtle variations in the rate of sample stirring, the sample temperature may differ from the calibration temperature, the ISE electrode may drift, the timing of the millivolt reading relative to insertion of the ISE electrode (Hargrave et al., 2008b; Wong and Page, 2025a, b).

A millivolt meter commonly used in association with ISE measurements is the Accumet AP125. This meter was used by Chang et al. (2014) and has a manufacturer's accuracy specification of ±0.2 mV (Fischer Scientific, 2014). The Canadian Federal Aquaculture Monitoring Standard associated with the Canadian Federal Aquaculture Activities Regulations require the use of a meter with a resolution of ± 0.1 mV (DFO, 2018). We consider this as a best estimate of the practical accuracy since an actual reading of mV is likely greater than this due to the factors mentioned above that either cause a systematic error in a mV reading or cause mV readings to "jump" around and requires the analyst to estimate a central tendency value. The practical accuracy may therefore be between about ± 0.2 mV and ± 1 mV or more.

Millivolt readings associated with ISE measurements of sulfide usually stabilize within the first 1-2 minutes after insertion of the ISE electrode into the sediment – SAOB slurry (Hargrave et al., 2008b).

The effect of error in the millivolt reading on the log transformed sulfide concentration is given by Eq. A1- 2 and an example is shown in Table A5-1. In the table the True mV column is the millivolt value corresponding to the true $\log_{10}$ sulfide value. The Upper and Lower limits of sulfide are obtained by adding and subtracting the specified error in millivolts to the true millivolts and solving for the $\log_{10}$ sulfide concentration. For millivolt errors of ±0.1 mV the corresponding $\log_{10}$(S) values have an absolute difference of 0.003 log units from the true value and a relative difference of less than 0.17%. For millivolt errors of ±1 mV the corresponding log10(S) values have an absolute difference of 0.034 log units and a relative difference of less than 1.7%. The table also illustrates that the absolute differences are independent of sulfide concentration and that the relative differences decrease with increasing concentration for a given millivolt uncertainty. In all cases the absolute and relative bias is zero.

$$V_S = \beta_0 + \beta_1 log_{10}(S_{NC}) + \varepsilon \qquad \text{Eq. A5-1}$$

$$log_{10}(S_{NC}) = (V_S \pm \varepsilon - \beta_0)/\beta_1 \qquad \text{Eq. A5-2}$$

Table A5-1: Effect of uncertainty in millivolt readings on log_{10} sulfide concentrations. Absolute uncertainty is calculated as half the difference between Upper and Lower concentrations. Relative uncertainty is the absolute uncertainty divided by the true concentration. Absolute bias is calculated as the average of the Upper and Lower concentrations minus the true concentration. Relative bias is the absolute bias divided by the true concentration. Abs – Absolute, Rel – Relative. The calculations assume a calibration relationship with intercept $\beta 0$ and slope $\beta 1$.

Error (mV)	True mV	Sulfide Concentration log_{10}(S) (*b1* = -29.5, *b0* = -797.9)			Uncertainty (Range/2)		Bias	
		True Value (µM)	Lower Limit (µM)	Upper Limit (µM)	Abs (± µM)	Rel (± %)	Abs (± µM)	Rel (± %)
0	-856.9	2.0	2.00	2.00	0.0	0.0	0.00	0.00
	-877.5	2.7	2.70	2.70	0.0	0.0	0.00	0.00
	-886.4	3.0	3.00	3.00	0.0	0.0	0.00	0.00
	-907.0	3.7	3.70	3.70	0.0	0.0	0.00	0.00
	-915.9	4.0	4.00	4.00	0.0	0.0	0.00	0.00
0.1	-856.9	2.0	2.00	2.00	0.000	0.00	0.00	0.00
	-877.5	2.7	2.70	2.70	0.000	0.00	0.00	0.00
	-886.4	3.0	3.00	3.00	0.000	0.00	0.00	0.00
	-907.0	3.7	3.70	3.70	0.000	0.00	0.00	0.00
	-915.9	4.0	4.00	4.00	0.000	0.00	0.00	0.00
1	-856.9	2.0	2.00	2.00	0.000	0.0	0.00	0.00
	-877.5	2.7	2.70	2.70	0.000	0.0	0.00	0.00
	-886.4	3.0	3.00	3.00	0.000	0.0	0.00	0.00
	-907.0	3.7	3.70	3.70	0.000	0.0	0.00	0.00
	-915.9	4.0	4.00	4.00	0.000	0.0	0.00	0.00
10	-856.9	2.0	2.00	2.00	0.000	0.0	0.00	0.00
	-877.5	2.7	2.70	2.70	0.000	0.0	0.00	0.00
	-886.4	3.0	3.00	3.00	0.000	0.0	0.00	0.00
	-907.0	3.7	3.70	3.70	0.000	0.0	0.00	0.00
	-915.9	4.0	4.00	4.00	0.000	0.0	0.00	0.00

The error in untransformed sulfide concentration generated by uncertainty in the mV reading is derived from Eq. A1- 3, which is the inverse of the calibration equation (Eq. A5- 2). The absolute limits are given by Eq. A5-1 to A5-4 and Eq. A5- 5 and the relative limits (U_{Srel}) are given by

Eq. A1- 6. In these equations the sulfide measurement with error is represented by S_M and the true sulfide concentration, i.e., the measurement without error ($\varepsilon = 0$) is represented by S_T.

An example set of calculations is given in Table A5-1. In this table the True mV column is the millivolt value corresponding to the true sulfide value. The Upper and Lower limits of sulfide are obtained by adding and subtracting the specified error in millivolts to the true millivolts and solving for the sulfide concentration. The absolute uncertainty in sulfide concentration generated by uncertainty in voltage measurements varies with the magnitude of the uncertainty in the millivolt readings whereas the relative uncertainty does not vary with concentration. For millivolt errors of ±0.1 mV the corresponding sulfide values have an absolute difference of 0.8 to 78 μM from the true value and a relative difference of 0.8%. For millivolt errors of ±1 mV the corresponding uncertainties are an order of magnitude larger than ±0.1 mV. The upper and lower sulfide values now have an absolute difference of 7.8 to 781 μM from the true value and a relative difference of 7.8%. This is the value indicated by Rundle (2011) as being typical for ISE measurements of divalent ions such as sulfide. Unlike the differences in transformed sulfide the absolute difference increases with concentration and the relative difference remains constant for a given millivolt uncertainty.

Also, unlike the transformed sulfide there is now a positive bias introduced by the non-linear i.e., logarithmic nature of the calibration relationship. This occurs because the nonlinear nature of the calibration relationship generates an asymmetry in the sulfide estimates relative to the true value. For a constant millivolt uncertainty, the absolute bias, defined as the average of the upper and lower concentrations associated with the millivolt uncertainty minus the true concentration, is seen to increase with the millivolt uncertainty and with the concentration (Table A5-2). The relative bias, defined as the absolute bias divided by the true concentration, also increases with the millivolt uncertainty and unlike the total uncertainty, the relative bias is independent of concentration. In the ±0.1 and ±1 mv cases the relative bias is small, i.e., less than 1%. However, as the error in millivolts increases the relative bias increases so that a millivolt error of 10 mV results in a positive bias of 32% in sulfide. The Aquaculture Monitoring Standard 2018 associated with the Canadian Aquaculture Activity Regulations requires the use of an ISE electrode and meter that achieves a 5% accuracy or better (DFO, 2018).

$$S_M = 10^{(V \pm \varepsilon - \beta_0)/\beta_1} \qquad \text{Eq. A5-3}$$

$$S_M^U = 10^{(V+\varepsilon-\beta_0)/\beta_1} = 10^{(V-\beta_0)/\beta_1} 10^{+\varepsilon/\beta_1} \qquad \text{Eq. A5-4}$$

$$S_M^L = 10^{(V-\varepsilon-\beta_0)/\beta_1} = 10^{(V-\beta_0)/\beta_1} 10^{-\varepsilon/\beta_1} \qquad \text{Eq. A5-5}$$

$$U_{Srel} = \frac{S_M}{S_T} = \frac{10^{(V \pm \varepsilon - \beta_0)/\beta_1}}{10^{(V-\beta_0)/\beta_1}} = 10^{\pm \varepsilon/\beta_1} \qquad \text{Eq. A5-6}$$

Unfortunately, the number of digits displayed by a millivolt meter may not be indicative of the true accuracy of the voltmeter and the analyst should investigate the accuracy of the meter more fully to better determine its accuracy (Taylor, 1997).

These features of uncertainty are illustrated in Figure A5-1 in which an uncertainty in millivolts was assumed to be normally distributed within a mean of zero mV and a standard deviation of one mV. Thirty sulfide concentrations were generated for five levels of millivolts that corresponded with five each standard sulfide concentrations (100, 500, 1,000, 5,000 and 10,000 μM). The calibration relationship assumed was representative of the calibration relationships reported in Chang et al. (2014).

Table A5-2: Effect of uncertainty in millivolt readings on untransformed sulfide concentrations. Absolute uncertainty is calculated as half the difference in Upper and Lower concentrations. Relative uncertainty is the absolute uncertainty divided by the true concentration. Absolute bias is calculated as the average of the Upper and Lower concentrations minus the true concentration. Relative bias is the absolute bias divided by the true concentration. Abs – Absolute, Rel – Relative. The calculations assume a calibration relationship with intercept $\beta 0$ and slope $\beta 1$.

Error (mV)	True mV	Sulfide Concentration (*b1* = -29.5, *b0* = -797.9)			Uncertainty (U) (Range/2)		Bias	
		True Sulfide (µM)	Lower Limit (µM)	Upper Limit (µM)	Abs (± µM)	Rel (± %)	Abs (± µM)	Rel (± %)
0	-856.9	100	100	100	0.0	0.0	0.00	0.000
	-877.5	500	500	500	0.0	0.0	0.00	0.000
	-886.4	1000	1000	1000	0.0	0.0	0.00	0.000
	-907.0	5000	5000	5000	0.0	0.0	0.00	0.000
	-915.9	10000	10000	10000	0.0	0.0	0.00	0.000
0.1	-856.9	100	99	101	0.8	0.8	0.00	0.003
	-877.5	500	496	504	3.9	0.8	0.02	0.003
	-886.4	1000	992	1008	7.8	0.8	0.03	0.003
	-907.0	5000	4961	5039	39.0	0.8	0.15	0.003
	-915.9	10000	9922	10078	78.1	0.8	0.30	0.003
1	-856.9	100	92	108	7.8	7.8	0.3	0.305
	-877.5	500	462	541	39.1	7.8	1.5	0.305
	-886.4	1000	925	1081	78.1	7.8	3.0	0.305
	-907.0	5000	4625	5406	390.7	7.8	15.2	0.305
	-915.9	10000	9249	10812	781.3	7.8	30.5	0.305
10	-856.9	100	46	218	86.2	86.2	32.0	32.0
	-877.5	500	229	1091	431.1	86.2	160.2	32.0
	-886.4	1000	458	2183	862.2	86.2	320.4	32.0
	-907.0	5000	2291	10913	4311.2	86.2	1602.0	32.0
	-915.9	10000	4582	21826	8622.4	86.2	3204.0	32.0

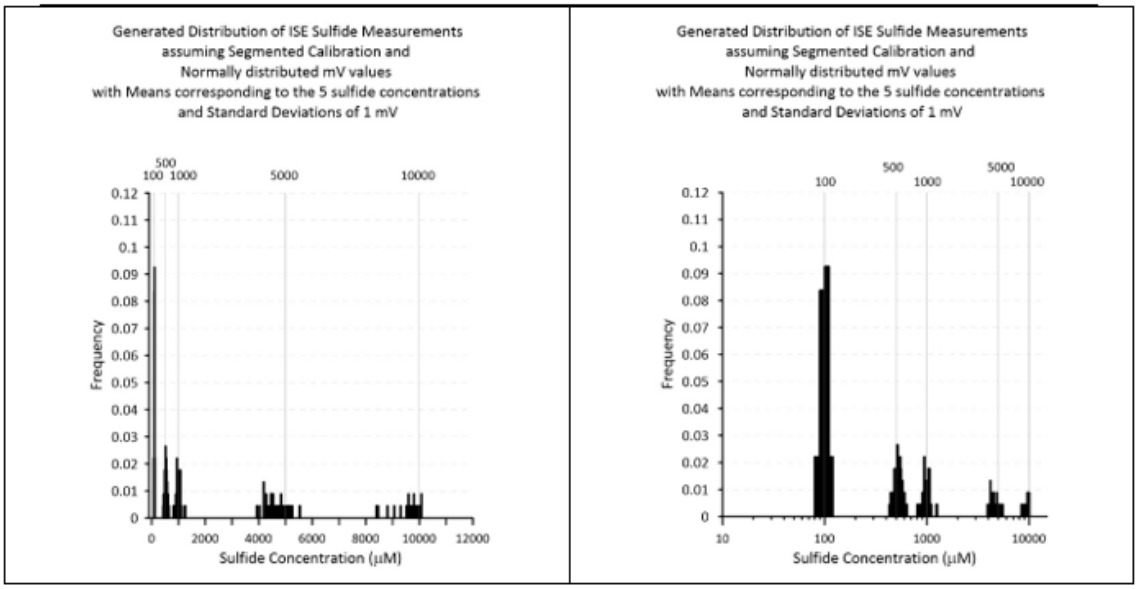

Figure A5-1: Simulated distributions of sulfide and log sulfide concentrations assuming a standard deviation of 1mV and specific mean concentrations of sulfide.

APPENDIX 6: PROPAGATION OF ERROR IN STANDARD CONCENTRATIONS

When only the error associated with the calibration standard solutions or nominal concentrations is considered the relative error is small. Uncertainty in the titration of standard solutions has been suggested to be ±0.1% (TFS, 2007). When only this error is considered the calibration equation becomes Eq. A6-1 and the sulfide concentration (S_{NC}) is given by Eq. A6-2. The relative error in S_{NC} is given by dividing the concentration that contains uncertainty by the concentration assuming no uncertainty i.e., $\alpha = 0$ (Eq. A6-3).

$$V_S = \beta_0 + \beta_1 log_{10}(S_{NC} \pm \epsilon) = \beta_0 + \beta_1 log_{10}(S_{NC} \pm \alpha S_{NC}) \quad \text{Eq. A6-1}$$

$$S_{NC} = \frac{10^{(V-\beta_0)/\beta_1}}{1+\alpha} \quad \text{Eq. A6-2}$$

$$U_{Srel} = \frac{\frac{10^{(V-\beta_0)/\beta_1}}{1+\alpha}}{10^{(V-\beta_0)/\beta_1}} = \frac{1}{1+\alpha} \quad \text{Eq. A6-3}$$

These equations indicate the uncertainty in ISE sulfide measurements generated by a titration error of 0.1% ($\alpha = 0.1/100 = 0.001$) is about 0.999 to 1.001. If this error is increased by a factor of ten to 1% ($\alpha = 1/100 = 0.01$) to account for dilution errors and perhaps less than ideal laboratory technique, the relative uncertainty increases to 0.99 to 1.01. This magnitude of error is similar to that associated with the accuracy of the voltage meter.

SEGMENTED CALIBRATION

The Accumet meter estimates the slope between decadal differences in calibration points and estimates sulfide concentrations for samples by applying a linear interpolation of the calibration relationship that contains the measured voltage and inverts the estimated value of log10(S).

In the segmented calibration approach the linear relationship between the Y variable (mV) and the X variable (log10 sulfide) described above is divided into segments with each segment corresponding to two adjacent calibration points. For example, in a three-point calibration there are three calibration coordinate pairs, (X1, Y1), (X2, Y2) and (X3, Y3) and two linear segments, one between (X1, Y1) and (X2, Y2) and another between (X2, Y2) and (X3, Y3). The basic linear relationship for each segment is the basic linear interpolation equation (Eq. A6-4) in which Y is the unknown mV associated with a known log sulfide (X).

$$Y = Y_1 + \frac{(Y_2 - Y_1)}{(X_2 - X_1)}(X - X_1) \quad \text{Eq. A6-4}$$

In practice, the inverse of Eq. A6- 4, i.e., Eq. A6- 5 and Eq. A6-6 are the functional relationships that relate the measured value (Y in mV) to the desired value (X in log sulfide).

$$X = X_1 + \frac{(X_2 - X_1)}{(Y_2 - Y_1)}(Y - Y_1) \quad \text{Eq. A6-5}$$

$$\log(S) = \log(S_1) + \frac{(\log(S_2) - \log(S_1))}{(mV_2 - mV_1)}(mV - mV_1) \quad \text{Eq. A6-67}$$

In this approach the errors associated with mV readings are assumed to cancel since they are error associated with the calibration mV readings (mV1, mV2) are assumed to be the same as the error associated with the mV reading on the unknown sample (mV). The only way to estimate the error in estimates of log(S) associated is to conduct some empirical experiments using measured values of mV and known values of log(S) and compare these to the estimated values using Eq. A6- 6.

APPENDIX 7: WILCOXON SIGNED-RANKS TEST FOR SMALL SAMPLE SIZES

Many research and regulatory sampling efforts collect a small number of samples, due in part to cost and time considerations, and the sample values are not necessarily normally distributed. This can limit the ability to make strong probabilistic inferences from the data. Sampling sediments and making measurements of pore water sulfide concentrations is not an exception.

The null hypothesis considered is that the sample measurements come from a population of measurements that have a median sulfide equal to a specified median (H0: Smed = Sthr). The scenarios only consider the alternative hypothesis that the sample median is greater than the specified threshold (HA: Smed > Sthr). When the null hypothesis is true, there is a certain probability that a sample drawn from the population will have a median value greater than the null distribution. When the probability is small, traditionally when the probability (p) is less than 5% (p<0.05), the inference is that the sample is unlikely to have come from the null distribution, and the sample is inferred to be significantly different from the null distribution. However, there is nothing magic about the 5%, the regulator can choose any probability that it deems suitable for its purposes.

The statistical test used is the non-parametric Wilcoxon signed ranks test. This statistic was chosen because the sample measurements may not be normally distributed and the sample sizes are typically small. In this test the test statistics and probabilities vary only in relation to relative rankings of the sample deviations from the regulatory threshold and the number of samples. For example, the sulfide concentrations 5,998, 5,999, and 6,005 μM results in the same test result as 4,500, 5,000, and 8,000 μM when the threshold of interest is 6,000 μM.

There are two scenarios considered. In one scenario the regulator is interested in whether the sample median is greater than a specified threshold. In the other scenario, the regulator is interested in whether the median of the sample population is greater than the median of the unimpacted or background population by a specified amount. The test statistic used is the Wilcoxon signed ranks test since the sample size is small, and the distribution of deviations from the threshold value is likely to be non-normal. In all scenarios the samples are assumed to be randomly selected and the measurements, and their deviations, are assumed to be symmetrically distributed about a median.

Only sample sizes of three, four, and five are shown.

SCENARIO ONE

The scenario one sample possible combinations are summarized in Tables A7-1 for sample sizes of three, four and five. In each table the possible combinations of sample ranks are shown along with the corresponding Wilcoxon test statistic and the probability the statistic value occurs when the null hypothesis is true. Only the ranks of the deviations of individual sample measurements from the threshold are shown because, as indicated above, it is only the relative combination of ranks that influences the outcome of the Wilcoxon sign ranks test, i.e., the scenario applies to all actual sample values that result in the indicated ranks.

The first table (Table A7-1) assumes three randomly selected measurements are available, that each measurement has had the threshold value subtracted from it and the signed rank of each value has been determined as per the Wilcoxon signed ranks test protocol (Sheskin, 2007). The nine possible combinations of ranked values, the Wilcoxon test statistics and the probability of the test statistic having the value it has under the assumption the null hypothesis is true are shown in. Only three of the nine possibilities give a qualitative suggestion that the sample may indicate the sampled population median may exceed the threshold and in all of these scenarios the largest difference between the measured value and the threshold is positive (S > Sreg) and

has a rank of +3. However, despite these suggestions, the test indicates the chance that any of the suggested threshold exceedances are solidly indicative of the sampled population actually having a median that this greater than the threshold is low. This is the same as saying the chance that the sampled population is actually from a population that has a median equal to the threshold value is quite large; the smallest chance or probability (*p*) given the sampling effort of n=3 is 0.125 (12.5%). In other words, decisions made on a sample size of three must be a) tolerant of an inability to detect exceedances of the desired threshold with little chance of being wrong, b) tolerant of an inability to detect deviations from thresholds at traditional *p* values of 0.05 or less, and c) tolerant of the fact the sampling effort will never detect exceedances if p values are to be less than 0.05 or even 0.1.

The second table (Table A7-2) assumes four randomly selected measurements are available (Table A4). In this situation there are thirteen combinations of ranks. Even when all samples have positive ranks, i.e., all measurements are greater than the threshold, the probability of this being true under the null hypothesis is only 0.0625 (6.25%).

The third table (Table A7-3) assumes five randomly selected measurements are available (Table A5). In this situation there are 23 combinations of ranks. All combinations have probabilities greater than 0.05 except the combination in which all ranks are positive, i.e., all measurements are greater than the threshold. The probability of this being true under the null hypothesis is 0.0313 (3.13%).

Table A7-1: Wilcoxon test statistics and one-sided p values associated with combinations of ranked differences between sample values and a constant threshold value for the case when n = 3. The probabilities (p) have been determined using the Wilcoxon signed ranks test function distributed with R version 4.1.1 (2021-08-10).

				H0: Smd = Sreg; HA: Smd > Sreg				
				Test Statistics			Interpretation	
	Signed Ranks of Absolute Differences			n = 3 Texp = 3			Qualitative Consistency with HA	Probability (*p* 1 sided)
Comment	#1	#2	#3	W^+	W^-	T	(yes or no)	
S range includes Sreg								
one S > Sreg	-1	-2	+3	3	3	3	no	0.625
one S > Sreg	-1	+2	-3	2	4	2	no	0.750
one S > Sreg	+1	-2	-3	1	5	1	no	0.875
two S > Sreg	-1	+2	+3	5	1	1	yes	0.250
two S > Sreg	+1	-2	+3	4	2	2	yes	0.500
two S > Sreg	+1	+2	-3	3	3	3	no	0.625

Comment	Signed Ranks of Absolute Differences			H0: Smd = Sreg; HA: Smd > Sreg				
				Test Statistics			Interpretation	
				n = 3 Texp = 3			Qualitative Consistency with HA (yes or no)	Probability (*p* 1 sided)
	#1	#2	#3	W^+	W^-	T		
S range does not include Sreg								
three S < Sreg	+1	+2	+3	6	0	0	yes	0.125
three S < Sreg	-1	-2	-3	0	6	0	no	1.000

Table A7-2: Wilcoxon test statistics and one-sided p values associated with combinations of ranked differences between sample values and a constant threshold value for the case when n = 4. The probabilities (p) have been determined using the Wilcoxon signed ranks test function distributed with R version 4.1.1 (2021-08-10).

Comment	Signed Ranks of Absolute Differences				H0: Smd = Sreg; HA: Smd > Sreg				
					Test Statistics			Interpretation	
					n = 4 Texp = 5			Qualitative Consistency with HA (yes or no)	Probability (*p* 1 sided)
	#1	#2	#3	#4	W^+	W^-	T		
S range includes Sreg									
one S > Sreg	-1	-2	-3	+4	4	6	4	no	0.688
one S > Sreg	-1	-2	+3	-4	3	7	3	no	0.812
one S > Sreg	-1	+2	-3	-4	2	8	2	no	0.875
one S > Sreg	+1	-2	-3	-4	1	9	1	no	0.938
two S > Sreg	-1	-2	+3	+4	7	3	3	yes	0.312
two S > Sreg	-1	+2	-3	+4	6	4	4	yes	0.438
two S > Sreg	+1	-2	-3	+4	5	5	5	no	0.562

Comment	Signed Ranks of Absolute Differences				H0: Smd = Sreg; HA: Smd > Sreg				
					Test Statistics			Interpretation	
					n = 4 Texp = 5			Qualitative Consistency with HA (yes or no)	Probability (*p* 1 sided)
	#1	#2	#3	#4	W^+	W^-	T		
two S > Sreg	-1	+2	+3	-4	5	5	5	no	0.562
two S > Sreg	+1	+2	-3	-4	3	7	3	no	0.812
three S > Sreg	-1	+2	+3	+4	9	1	1	yes	0.125
three S > Sreg	+1	+2	+3	-4	6	4	4	yes	0.438
S range does not include Sreg									
four S > Sreg	+1	+2	+3	+4	10	0	0	yes	0.062
four S < Sreg	-1	-2	-3	-4	0	10	0	no	1.000

Table A7-3: Wilcoxon test statistics and one-sided p values associated with combinations of ranked differences between sample values and a constant threshold value for the case when n = 5. The probabilities (p) have been determined using the Wilcoxon signed ranks test function distributed with R version 4.1.1 (2021-08-10).

Comment	Signed Ranks of Absolute Differences					H0: Smd = Sreg; HA: Smd > Sreg				
						Test Statistics			Interpretation	
						n = 4 Texp = 5			Qualitative Consistency with HA (yes or no)	Probability (*p* 1 sided)
	#1	#2	#3	#4	#5	W^+	W^-	T		
S range includes Sreg										
one S > Sreg	-1	-2	-3	-4	+5	5	10	5	no	0.781
one S > Sreg	-1	-2	-3	+4	-5	4	11	4	no	0.844
one S > Sreg	-1	-2	+3	-4	-5	3	12	3	no	0.906

Comment	Signed Ranks of Absolute Differences					H0: Smd = Sreg; HA: Smd > Sreg				
						Test Statistics			Interpretation	
						n = 4 Texp = 5			Qualitative Consistency with HA (yes or no)	Probability (*p* 1 sided)
	#1	#2	#3	#4	#5	W^+	W^-	T		
one S > Sreg	-1	+2	-3	-4	-5	2	13	2	no	0.938
one S > Sreg	+1	-2	-3	-4	-5	1	14	1	no	0.969
two S > Sreg	-1	-2	-3	+4	+5	9	6	6	yes	0.406
two S > Sreg	-1	-2	+3	-4	+5	8	7	7	yes	0.500
two S > Sreg	-1	+2	-3	-4	+5	7	8	7	no	0.594
two S > Sreg	-1	-2	+3	+4	-5	7	8	7	no	0.594
two S > Sreg	+1	-2	-3	-4	+5	6	9	6	no	0.688
two S > Sreg	-1	+2	-3	+4	-5	6	9	6	no	0.688
two S > Sreg	+1	-2	-3	+4	-5	5	10	5	no	0.781
two S > Sreg	-1	+2	+3	-4	-5	5	10	5	no	0.781
two S > Sreg	+1	+2	-3	-4	-5	3	12	3	no	0.906
three S > Sreg	-1	-2	+3	+4	+5	12	3	3	yes	0.156
three S > Sreg	-1	+2	+3	-4	+5	10	5	5	yes	0.312
three S > Sreg	-1	+2	+3	+4	-5	9	6	6	yes	0.406
three S > Sreg	+1	+2	-3	-4	+5	8	7	7	yes	0.500
three S > Sreg	+1	+2	+3	-4	-5	6	9	6	no	0.688

Comment	Signed Ranks of Absolute Differences					H0: Smd = Sreg; HA: Smd > Sreg				
						Test Statistics			Interpretation	
						n = 4 Texp = 5			Qualitative Consistency with HA (yes or no)	Probability (*p* 1 sided)
	#1	#2	#3	#4	#5	W^+	W^-	T		
four S > Sreg	-1	+2	+3	+4	+5	14	1	1	yes	0.062
four S < Sreg	+1	+2	+3	+4	-5	10	5	5	yes	0.312
S range does not include Sreg										
five S > Sreg	+1	+2	+3	+4	+5	15	0	0	yes	0.031
five S > Sreg	-1	-2	-3	-4	-5	0	15	0	no	1.000