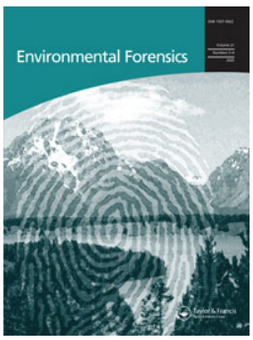

Environmental Forensics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uenf20

A review of marine water sampling methods for trace metals

A. E. Noble, C. B. Tuit, J. P. Maney & A. D. Wait

To cite this article: A. E. Noble, C. B. Tuit, J. P. Maney & A. D. Wait (2020) A review of marine water sampling methods for trace metals, Environmental Forensics, 21:3-4, 267-290, DOI: 10.1080/15275922.2020.1771629

To link to this article: https://doi.org/10.1080/15275922.2020.1771629

ENVIRONMENTAL FORENSICS
2020, VOL. 21, NOS. 3–4, 267–290
https://doi.org/10.1080/15275922.2020.1771629

Taylor & Francis
Taylor & Francis Group

A review of marine water sampling methods for trace metals

A. E. Noble[a,b], C. B. Tuit[a], J. P. Maney[a], and A. D. Wait[a]

[a]Gradient, Boston, MA, USA; [b]California Department of Toxic Substances Control, Sacramento, CA, USA

ABSTRACT

This review summarizes government agency and scientific organization guidance for collecting representative, reliable, and defensible water samples for trace metal analysis, and provides a framework for choosing appropriate sampling techniques to meet a study's data quality objectives, with respect to Natural Resource Damage Assessment in the marine environment. Preventing or limiting contamination is a primary concern during the sampling process, with the goal of constraining sampling-related contamination to below the targeted screening levels. Best practices should include clean-hands/dirty-hands sampling techniques and acid-washed plastic sample bottles. Quality control samples should be collected to detect and quantify contamination and to monitor the precision of the measurement process.

KEYWORDS

Natural resource damages; representative sampling; trace metals; marine water; quality assurance

Introduction

Natural Resource Damage (NRD) Assessment is the evaluation of injury to the environment and the determination of appropriate actions for restoring the environment to conditions as they were prior to that injury, which includes encouraging public involvement in that process (NOAA, 1996, pdf p. 12). Collecting reliable, defensible samples for assessing ecological risk/environmental injury and/or identifying sources of contamination is critical to the success of an NRD Assessment. Herein, we review current guidance from government agencies and scientific organizations and summarize sampling techniques appropriate for measuring trace metal concentrations in marine water with respect to NRD. As discussed in Wait et al. (2020), this issue, data are only as good as the weakest link in the data acquisition process, a process that starts with determining the measurements to be made. Guidance for the general collection of water samples is abundant and generally consistent across sources, but variability arises among guidance for collecting water samples for trace metal analysis. We summarize those parts of the guidance about which there is consensus and provide further guidance on navigating the remaining variability among different sampling approaches. We also present a guidance framework for choosing a sampling approach for trace metal sampling in marine waters (Figure 1), which is intended to help personnel interested in NRD Assessment select appropriate sampling protocols for their trace metal data quality objectives (DQOs). Though the focus of this paper is on the marine environment, consistent with the other papers in this series (Wait et al., 2020; Tuit et al., 2020; Kneeland et al., 2020; Tcaciuc et al., 2020 *Environmental Forensics*, this issue), many of the issues and concerns addressed in the paper are also applicable to freshwater systems.

In this review, we present a framework for recommended best practices in NRD Assessment trace metal sampling (Figure 1). Using this framework in conjunction with US EPA's NRALWQ Criteria (US EPA, 2017) as example screening levels for metals (*i.e.*, as example DQOs), we address guidance regarding contamination prevention (sample handling procedures, sampling equipment, sample bottles), the sampling procedure, sample processing (filtration and preservation), and QC samples (blanks, replicates, matrix spikes, *etc.*). This review is one of a series of 5 articles that focus on accepted sampling practices for NRD Assessment. Wait et al. (2020), in this issue, presents an introduction to forensic sampling practices with respect to oil spills in the marine environment, and the remaining three reviews, also in this issue, describe accepted practices for sampling hydrocarbons in the water column (Kneeland et al., 2020), sediments (Tuit et al., 2020), and oils (Tcaciuc et al., 2020).

CONTACT A. Dallas Wait dwait@gradientcorp.com Gradient, One Beacon Street, 17th Floor, Boston, MA 02108, USA.

Supplemental data for this article is available online at https://doi.org/10.1080/15275922.2020.1771629.
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

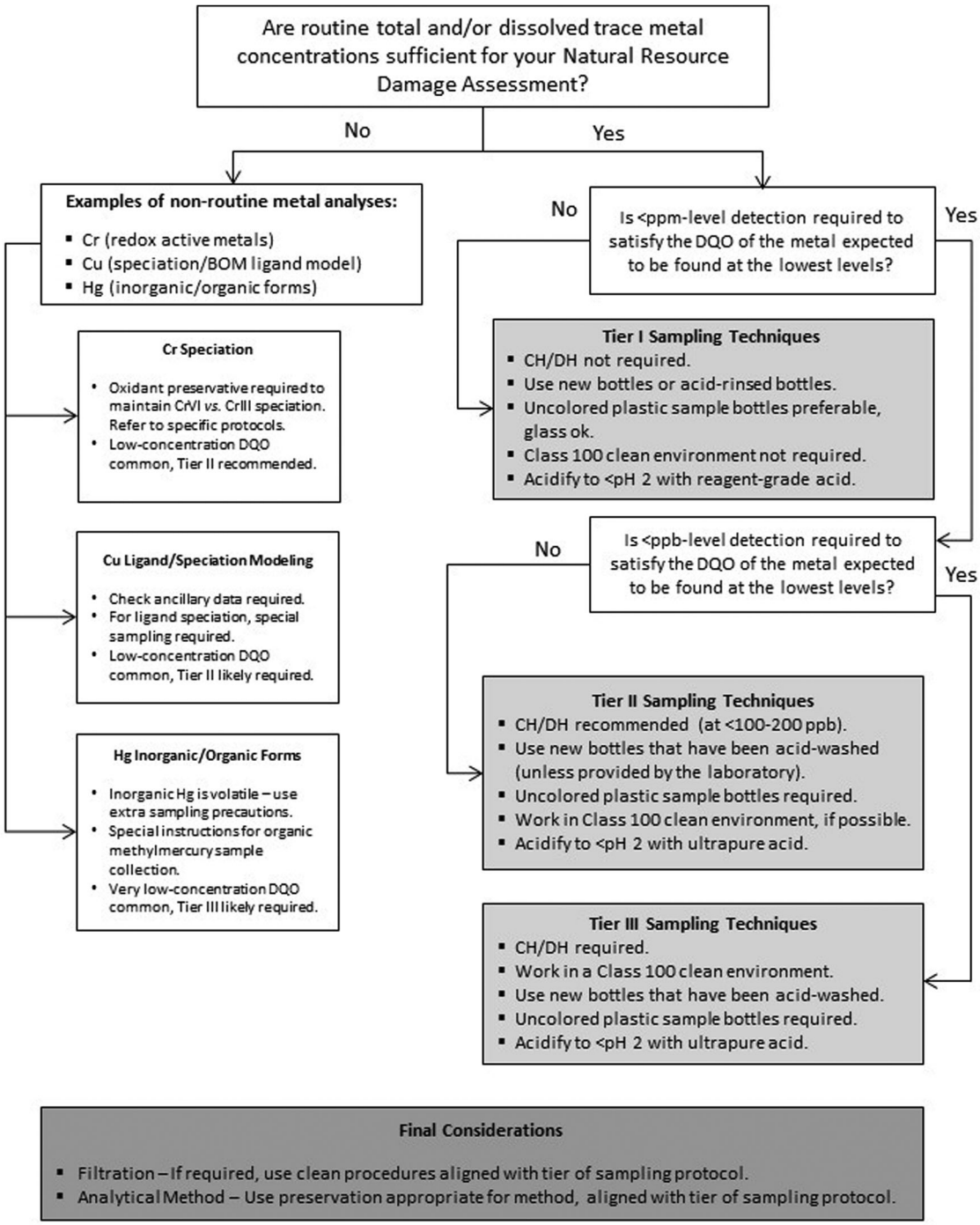

Figure 1. Guidance framework for choosing a trace metal sampling approach that will satisfy data quality objectives and ensure that representative samples can be successfully collected.

For additional information on the reviewed guidance, a comprehensive summary table of government, state, academic, and consensus standard guidance can be found in the Supplemental Materials (Supplemental Table 1). Supplemental Table 1 summarizes accepted practices for sample handling, sample bottles, equipment, preservation, holding times, and QC samples for different types of aqueous samples, and provides a cursory comparison that highlights the differences and similarities among them.

Establishing data quality objectives for trace metal sample collection

Before developing a sampling protocol, three initial decisions need to be made in order to define the

metrics by which an appropriate protocol can be developed. The choice of target analytes and detection limits dramatically influences the cost and effort required to carry out a successful sampling campaign. Figure 1 presents a framework for recommended best practices in NRD Assessment trace metal sampling that aligns the answers to these decisions with a tiered approach for successful sample collection. This tiered approach helps ensure that (a) the samples collected will be meaningful for achieving the DQOs and (b) that the samples are collected in a cost-appropriate manner. The framework is designed to be a key reference throughout the sample collection planning process.

The first decision is to select a suite of metal analytes that will be appropriate and meaningful for the assessment. This includes not only which metals should be analyzed, but also the form of those metals. Metals can be measured in water and used to identify injury to the environment or as forensic markers in several ways. Routine monitoring protocols that include a suite of trace metal analyses often call for the collection of "total" or "total recoverable" metal samples, in which whole water is collected and preserved. Total recoverable analyses are often the default if total *vs*. filtered analyses are not specified in a sampling protocol (ORDEQ, 2013, p. 15). Water sampling guidance often designates the collection of filtered samples as secondary to the collection of total samples (*e.g.*, "If dissolved metals and/or dissolved nutrients analysis is required…" [UTDEQ, 2011, emphasis added]), and while total metals analyses may be sufficient to satisfy a study's DQOs, there are often cases in which additional sample types are relevant or even critical for achieving the DQOs. For example, the US EPA National Recommended Aquatic Life Water Quality (NRALWQ) Criteria (US EPA, 2017) are based on dissolved metal concentrations (US EPA, 2004), rather than total metal concentrations. US EPA provides conversion factors (albeit, with limitations) from total concentrations to dissolved concentrations if only total metals samples are collected, but it may instead be advisable to collect samples for both total and filtered analyses.

Additionally, the speciation of a given metal may also influence the concentration at which it presents an identifiable risk. Examples of this include redox state (important for trivalent/hexavalent chromium [CrIII/VI] risk), inorganic/organic form (elemental [Hg] *vs*. monomethyl mercury [MMHg]), and chemical speciation (free copper ions [Cu'] *vs*. ligand-bound Cu [CuL]). It is important to establish whether successfully determining NRD with respect to the metal of interest is dependent upon one of these more-involved analytical determinations and whether this would subsequently affect the recommended sampling approach. For example, US EPA offers a biotic ligand model (BLM) that can estimate Cu chemical speciation from a dissolved Cu measurement, which alleviates the need for specialized Cu sampling but requires taking ancillary samples for other parameters (*e.g.*, total organic carbon, pH, salinity) to support the model calculations (US EPA, 2016a). While most metals can be preserved well in acid, inorganic Hg is volatile as a dissolved neutral species and requires that extra precautions be taken during sample preservation. Differentiating the organic *vs*. inorganic forms of Hg may also be important, because these forms present different levels of ecological risk, and they need to be preserved and sampled differently (US EPA, 2009, p. 160). While metal speciation sampling practices are important for some metals, these practices are beyond the scope of this review, which will focus on general trace metal sample collection and preservation practices intended for the analysis of total and dissolved metals.

The second major decision is to determine the DQOs' detection level requirements. DQO detection level requirements will depend on the targeted screening levels, the expected level of contamination, and expected background concentrations, if the DQOs require background measurements. It is important to determine the targeted screening levels for all metals of interest and the answers will determine the appropriate sampling methods as well as viable analytical techniques to use. For trace metals, the level of effort (and thus cost) required for contamination prevention when sampling at a screening level *versus* a background concentration level can vary significantly. This variance stems from two facts. First, metal concentrations in aqueous media can span several orders of magnitude, from sub-part per trillion (ppt) to >100 parts per million (ppm). For example, background iron (Fe) concentrations in open ocean seawater are on the order of ~0.004-0.1 part per billion (ppb) (Johnson et al., 1997), the chronic freshwater criteria established by the US EPA for Fe is 1,000 ppb (US EPA, 2017), and contaminating Fe concentrations can reach >100,000 ppb (USGS, 2012). Second, concentrations that cause NRD vary from metal to metal and also span several orders of magnitude. For example, Fe and Hg chronic freshwater criteria levels differ by 4 orders of magnitude (Fe = 1,000 ppb and Hg = 0.77 ppb; US EPA, 2017). As a result, it is important to tailor the sample collection approach to ensure successful sampling of the contaminant with the lowest

criteria value. For many metals, background concentrations in marine waters are typically lower than those in terrestrial surface water bodies, which are the focus of most government and state guidance on water sampling. Thus, it is very important to understand what metals and what screening or background levels are to be targeted, to ensure that your sampling practices are aligned with your DQOs. Again, for the purpose of this review, US EPA's NRALWQ Criteria (US EPA, 2017) are used to follow the framework. These criteria can be found in Table 1.

The third decision that should be made prior to establishing a sampling protocol is to select the analytical methods to be used for determining the trace metal concentrations in the collected samples. The analytical method can impact sampling parameters, such as sample volume and preservation, and the analytical method selected for a study must always have a method detection limit that is lower than the lowest study DQO, whether it is set at the criteria, screening, or background level. This may seem like an obvious statement, but it is not unusual to encounter datasets, either historical or those collected during the initial remedial investigation phases, with detection limits for one or more of the metals of interest that are above ecological risk or NRD Assessment criteria. Such data are unusable and a waste of resources. The analytical method detection limit is a performance-based value, meaning that it is determined at the time of analysis from the analytical blank of the method, it is laboratory- and metal-specific, and can vary from sample to sample. A discussion of analytical methodologies is beyond the scope of this review, but guidelines for determining an analytical method detection limit can be found in the "Report of the Federal Advisory Committee on Detection and Quantitation Approaches and Uses in Clean Water Act Programs (Final)" (Federal Advisory Committee on Detection and Quantitation Approaches and Uses in Clean Water Act Programs, 2007) and "Definition and Procedure for the Determination of the Method Detection Limit, Revision 2" (US EPA, 2016b).

After making decisions about these three DQO-driven considerations (types of samples to be collected, detection level requirements, and anticipated analytical techniques), environmental assessors should be able to select an appropriate sampling protocol for a NRD Assessment. An appropriate sampling protocol, including sample collection, processing (*e.g.*, filtration preservation), transportation, and storage, is also performance-driven. There is the potential for sample contamination and bias associated with each step in the sampling protocol, which must be documented with appropriate quality control (QC) samples, including, but not limited to, field equipment blanks, filter blanks, trip blanks, bottle blanks, and matrix spikes. This review will discuss several precautions to take and procedures for limiting contamination. Ideally, concentrations measured in the sampling blanks will be below the method detection limit of the analytical method selected, but in reality, they are often higher. Reported trace metal results should generally include the sampling blank concentration or the method detection limit, whichever is higher. Thus, results below these cutoff values should be reported as $< X$, where X is the sampling blank concentration or the method detection limit, whichever is higher. Similar to the method detection limit, if the sampling blank concentration for a metal exceeds its criteria or screening level, the resulting data will yield a false positive and may be considered unusable. It is not considered best practice to blank correct results for ecological risk or NRD Assessment, because the blank-corrected results could be underprotective. If the sampling blank concentration for a metal is above its criteria or screening level, it will be necessary to resample using more rigorously "clean" methods. Again, the framework shown in Figure 1, in conjunction with answers to the three decision points above, is designed to help environmental assessors select the level of effort that will produce reliable and representative sampling data without the necessity for costly and time-consuming resampling.

While it may seem safest to always employ the most stringent sampling techniques available, it may not always be necessary to put in this extra effort or expense. The sampling blank reporting limit for an analyte may be orders of magnitude higher than the true concentration or the detection capabilities of the analytical method, and this may not matter if the criteria or screening level for a metal are substantially higher than the sampling blank reporting limit. For example, some analytical techniques for determining Fe concentrations are capable of detecting concentrations below ~0.001 ppb, and background concentrations of Fe in the open ocean may approach this level. However, the Fe chronic freshwater criteria established by US EPA is 1,000 ppb, and an acceptable sampling blank to evaluate against this criterion could be as high as 100 ppb, depending on the DQOs. Even if the true Fe concentration in a sample may be significantly lower than 100 ppb, and the analytical technique used may be able to measure sub-ppb levels, the accuracy of the reported Fe concentration is limited by

Table 1. US EPA National Recommended Aquatic Life Water Quality Criteria and accompanying sample handling recommendations.

		EPA National Recommended Aquatic Life Water Quality Criteria						
		Freshwater[1]		Saltwater		Concentrations at Which CH/DH[2] Sample Handling Techniques are:		
Metal	Ambient Seawater Typical Concentration (μg/L)	CMC[3] Acute (μg/L)	CCC[4] Chronic (μg/L)	CMC[3] Acute (μg/L)	CCC[4] Chronic (μg/L)	Recommended[5]	Required[5]	Insufficient[6]
Aluminum	0.03	750	87	–	–		≤200 μg/L	
Arsenic	1.5	340	150	69	36	≤100 μg/L	~1 μg/L	
Cadmium	0.07	1.8	0.72	33	7.9	≤100 μg/L	~1 μg/L	
Chromium III	0.21	570	74			≤100 μg/L	~1 μg/L	
Chromium VI	0.21	16	11	1,100	50	≤100 μg/L	~1 μg/L	
Copper[7]	0.16	–	–	4.8 [2.1]	3.1 [1.3]	≤100 μg/L	~1 μg/L	
Iron	0.03	–	1,000	–	–		≤200 μg/L	
Lead	0.003	65	2.5	210	8.1	≤100 μg/L	~1 μg/L	
Mercury	0.0001	1.4	0.77	1.8	0.94	≤100 μg/L	~1 μg/L	<1 μg/L
Nickel	0.49	470	52	74	8.2	≤100 μg/L	~1 μg/L	
Selenium[8]	0.15	–	1.5 (lentic) 3.1 (lotic)	290	71	≤100 μg/L	~1 μg/L	
Zinc	0.36	120	120	90	81	≤100 μg/L	~1 μg/L	

Notes:

Sources: Ambient seawater concentrations = Monterey Bay Aquarium Research Institute (MBARI, 2017); US EPA National Recommended Aquatic Life Water Quality Criteria = US EPA (2017).

(1) Freshwater criteria are included for comparison, and because several trace metals have no saltwater criteria.

(2) Clean Hands/Dirty Hands. Bottle washing has been performed to standard, and equipment cleaning protocols are sufficient.

(3) Criterion maximum concentration.

(4) Criterion continuous concentration.

(5) Per US Geological Survey (USGS) *National Field Manual for the Collection of Water Quality Data* (USGS, 2015).

(6) May require more rigorous precautions, and clean environments such as glove boxes or Class 100 clean laboratories are required for processing samples in the field.

(7) 2016 US EPA-recommended values currently under consideration (updated criteria in brackets) (US EPA, 2016a).

(8) Selenium freshwater values are stipulated as not to be exceeded as a monthly average concentration in flowing (lotic) and standing (lentic) waters. These values are derived from a fish tissue value and a bio-accumulation model (US EPA, 2016c).

the value and reproducibility of the sampling blank. While extreme caution may be required to measure sub-ppb concentrations of some metals, this extra effort may be unnecessary for assessing Fe at the ppm level.

Contamination detection and prevention

The possible sources of metal contamination to a water sample are ubiquitous. Being aware of contamination sources, collecting blanks that will appropriately quantify contamination, and making efforts to minimize contamination are critical to the collection of representative, reliable, and defensible samples. As the sensitivity of modern instrumentation has increased, the impact of contamination on analytical results has become increasingly significant. In 1957, R.E. Thiers identified two different types of contamination in trace metal analysis: "negative" and "positive" contamination (Thiers, 1957). These two terms were later defined by US EPA (1986, pp. 9–46) guidance as:

1. Negative contamination: The potential for the measured analyte concentration to be artificially low because of losses from volatilization or adsorption.
2. Positive contamination: The potential for the measured analyte to be artificially high because of leaching or the introduction of foreign matter into the sample by particle fallout or gaseous air contaminants.

By 1975, Karin et al. (1975) had confirmed the work of earlier researchers that, among other types of containers, polyethylene sample containers were a source of positive trace metal contamination and that leaching containers with nitric acid prior to use was an efficient means of preventing potential contamination from this source. As early as 1976, it was known that contamination was causing the misreporting of lead concentrations in waters and that appropriate sampling and analysis lowered those reported concentrations by several orders of magnitude (Patterson and Settle, 1976). Thus, the initial focus in trace metal analysis was on positive contamination, and reported trace metal concentrations decreased as advances in clean sampling approaches (*e.g.*, use of clean benches, clean rooms, and improved leaching and sampling procedures) were introduced and adopted. However, at this time, researchers were also concerned with the potential for negative contamination caused by active adsorptive sites on sampling equipment and sample containers (Litman et al., 1975; Patterson et al., 1976). We address negative contamination further in our discussion of sample bottles, below.

Concerns regarding negative and positive contamination are just as pertinent today. The following notable excerpts from US federal guidance help underscore the importance of contamination prevention for trace metal analysis:

> Preventing ambient water samples from becoming contaminated during the sampling and analytical process is the greatest challenge faced in trace metals determinations. In recent years, it has been shown that much of the historical trace metals data collected in ambient water are erroneously high because the concentrations reflect contamination from sampling and analysis rather than ambient levels (Reference 12). Therefore, it is imperative that extreme care be taken to avoid contamination when collecting and analyzing ambient water samples for trace metals. (US EPA, 1996a)

> Think contamination! To ensure the integrity of the sample, be aware of possible sources of contamination. Contamination introduced during each phase of sample collection (and processing) is additive and usually is substantially greater than contamination introduced elsewhere in the sample-handling and -analysis process. Therefore, collect a sufficient number of quality-control samples, appropriately distributed in time and space, to ensure that data-quality objectives and requirements are met. (USGS, 2015, emphasis in original)

The importance of collecting sufficient QC samples and employing good contamination prevention measures cannot be overstated. To achieve both, it is first important to understand what is needed to carry out all aspects of sampling in a manner that will comply with a study's DQOs. At minimum, one must be able to successfully quantify any contamination introduced during the entire sample collection, handling, and preservation process. This is usually achieved with the collection and analysis of one or more types of blanks (*e.g.*, container, bottle, equipment, field, trip). The contaminant concentration in blanks must be sufficiently below the target screening levels (and sufficiency should be dictated by the DQO for each contaminant). In most cases, achieving this requires three key components: "clean hands/dirty hands" (CH/DH) sample handling procedures, acid-washed plastic sampling equipment, and acid-washed plastic sample bottles. For each of these components, key decisions must be made regarding the necessary precautions and QC sufficient to satisfy the study's DQOs. Awareness of potential sources of contamination is a good place to start when identifying the number and types of blanks necessary to achieve a study's DQOs.

Table 2 provides an illustrative, but not exhaustive, list of components of the sampling process, the associated sources of potential contamination, and the types of QC samples that can be collected to identify whether any of these potential sources are indeed a source of contamination in a given sampling program. This list can be used as a guide to select appropriate blanks to meet DQOs.

Guidance on how to navigate these issues during the entire sampling process is laid out in the following sections, using US EPA's NRALWQ Criteria (US EPA, 2017) as the example DQOs. For other DQOs, be they more or less stringent than the NRALWQ Criteria, the framework (Figure 1) provides general guidance for when the sampling procedures required may deviate from those discussed here. In general, NRALWQ Criteria will require Tier II sampling techniques, which are generally acceptable for trace metal concentrations within the 1-200 ppb range. Tier I sampling techniques are less laborious and costly than Tier II techniques (which are appropriate for measuring ppm-level concentrations and sufficient for some metals at concentrations >200 ppb), and Tier III sampling techniques (which are appropriate for measuring sub-ppb-level concentrations) are more laborious and costly.

Sample handling procedures (clean hands/dirty hands)

One of the most important decisions regarding proper contamination prevention is to choose a level of precaution for sample handling procedures that matches the defined DQOs. A cautious approach is recommended when beginning a sampling program in which the magnitude of the metal analyte concentration is unknown. This generally includes careful practice of a CH/DH sample handling technique, which is discussed in detail by many sampling guidance documents (US EPA, 1996a; VADEQ, 1996; FLDEP, 2014a; USGS, 2015; ADEQ, 2015). CH/DH sample handling is germane to Tier II and Tier III sampling techniques (Figure 1). The general purpose of the CH/DH sample handling technique is to ensure that only uncontaminated, clean plastic comes into contact with a sample at any given time. This technique is described in detail in US EPA Method 1669:

> Upon arrival at the sampling site, one member of the two-person sampling team is designated as "dirty hands"; the second member is designated as "clean hands." All operations involving contact with the sample bottle and transfer of the sample from the sample collection device to the sample bottle are handled by the individual designated as "clean hands." "Dirty hands" is responsible for preparation of the sampler (except the sample container itself), operation of any machinery, and for all other activities that do not involve direct contact with the sample …
>
> Sampling personnel must wear clean, nontalc gloves (Section 6.7) during all operations involving handling of the [sampling] Apparatus, samples, and blanks. Only clean gloves may touch the Apparatus. If another object or substance is touched, the glove(s)

Table 2. Components of trace metal sample collection processes and associated sources of contamination.

Components of Sample Collection	Sources of Contamination	QC Samples to Identify/Quantify Contamination Sources
Sampling equipment	Dust/airborne particles Metal materials Dirty parts Non-metal materials that may leach metals (*e.g.,* grease, rubber O-rings) Equipment storage materials	Equipment rinsate blank
Sample storage bottles	Dust/airborne particles Bottle material Equipment used to manufacture the bottle Extra bottle components (*e.g.,* cap liners, O-rings) Colored plastics	Bottle rinsate blank before cleaning Bottle rinsate blank after cleaning
Sample handling procedures	Dust/airborne particles Glove type (powdered) Unintentional glove contamination (*e.g.,* touching a door knob, brushing hair out of face) Sampling personnel (skin, hair, spit, clothing)	Field blank (leave a bottle open in the sampling area during sample collection)
Sample post-processing (filtration and preservation)	Dust/airborne particles Filters/filter cartridges/filter holders Acid used for preservation Equipment used for preservation or filtration (*e.g.,* pipette tips, forceps)	Filter blanks Equipment rinsate blanks Acid blanks and/or preservative blanks
Cleaning procedures (applies to all cleaning steps)	Water used for cleaning or rinsing Acid used for cleaning and rinsing Soaps used for cleaning and rinsing Other reagents used for cleaning and rinsing	Water blank Cleaning solution blanks

> must be changed before again handling the Apparatus. If it is even suspected that gloves have become contaminated, work must be halted, the contaminated gloves removed, and a new pair of clean gloves put on. Wearing multiple layers of clean gloves will allow the old pair to be quickly stripped with minimal disruption to the work activity. (US EPA, 1996a)

As noted in the above guidance excerpt, the CH/DH technique requires at least two sampling personnel to carry out. CH/DH requirements include the use of plastic, disposable, powder-free gloves. US EPA (1996a) specifies that the gloves be polyethylene, latex, vinyl, or polyvinyl chloride (PVC), and that "Shoulder-length gloves are needed if samples are to be collected by direct submersion of the sample bottle into the water or when sampling for mercury" (US EPA, 1996a, p. 14). One of the driving tenets of the CH/DH technique is that "Gloves should only contact surfaces that are metal-free. If there is any question as to whether the gloves are contaminated, change the gloves" (OWRB, 2003).

While these sample handling procedures are clearly laborious, they are necessary for collecting representative samples that can be analyzed for many metals at NRD Assessment criteria levels (though depending on the project DQOs, a less stringent sample handling procedure may be sufficient). Table 1 demonstrates how background concentrations of metals in the open ocean compare to US EPA's NRALWQ Criteria, and to the recommended CH/DH approach to sample handling. The decision to employ CH/DH methods is performance-driven. If contamination levels in sampling blanks are too high to meet the study DQOs, then stricter protocols should be enacted. The expected order of magnitude range of concentrations and detection requirements are critical pieces of information for selecting an appropriate sampling protocol. As can be seen in the framework (Figure 1), this could mean the difference between employing the above exacting and expensive sampling approach and choosing a less stringent sampling approach. Consistent with the potential for order-of-magnitude differences in the possible concentrations encountered at the study site, some sampling protocols explicitly specify using clean *vs.* "normal" protocols for some analytes. For example, the State of Georgia specifies that "Trace-metal detection sampling involves yet another level of equipment preparation, QA/QC documentation and anti-contamination measures" (GADNR, 2008), and provides information regarding specialized methods to use when sampling for trace metals (and thus CH/DH sample handling) is required. The Florida Department of Environmental Protection (FLDEP) summarizes this difference in their Standard Operating Procedure (SOP) for Clean Sampling for Ultratrace Metals in Surface waters, along with some caveats that may influence the decision to employ CH/DH procedures: "This method is not intended for determination of metals normally found in treated and untreated discharges from industrial facilities. Actual concentration ranges to which this guidance is applicable will depend on the sample matrix, dilution levels and other laboratory operating conditions" (FLDEP, 2014a). In addition, if the study's DQOs are driven by US EPA's NRALWQ Criteria, it is best to employ CH/DH techniques.

Implicit in these guidance documents is the necessity for personnel to be trained in sampling protocols beyond that used for typical environmental sampling. Even for typical environmental water sampling, government agency field sampling guidance commonly emphasizes that field sampling personnel must be properly trained. For example, the "Collection of Water Samples" chapter in the US Geological Survey (USGS) *National Field Manual for the Collection of Water-Quality Data* (USGS, 2015) states that "Formal training and field apprenticeship are necessary in order to implement correctly the procedures described in this chapter." According to US EPA and US Army Corps of Engineers (US ACE) guidance, "Sample collection requires an experienced crew, an adequate vessel equipped with navigational and supporting equipment appropriate to the site and the study, and noncontaminating sampling apparatus capable of obtaining relatively undisturbed and representative samples" (US EPA and US ACE, 1995). A trace metal sampling protocol should also include these requirements, at minimum.

Employing the appropriate level of contamination prevention can be particularly challenging when metals are not the primary analytes of concern in the study, if the order of magnitude of trace metal concentrations expected to be found in the study is unknown, or if there are several orders of magnitude difference in criteria concentrations among the metals slated for analysis. For example, the US EPA NRALWQ Criteria for zinc (Zn) is 81 μg/L, compared to 1.3 μg/L for Cu (Table 1), and the USGS (2015) sample handling recommendations are different for these concentrations. Using the CH/DH sample handling technique is only recommended at concentrations $\leq$100 μg/L, but is required when the target concentrations are in the single-digit μg/L range. Thus, it is important to know the low end of the concentration range predicted to be

present for each analyte and to adjust the study's sampling practices accordingly.

While CH/DH sample handling generally requires two people to perform properly, some guidance suggests that clean sampling can be accomplished by a single person, assuming they have the proper training and remain alert to sources of contamination:

> When sampling waters containing trace metals, nutrients or organic compounds, a single person wearing plastic disposable gloves and taking appropriate care can carry out the operation without sample contamination if he or she is alert to potential sources of contamination. Lavish use of polyethylene sheeting to wrap equipment and to cover work areas on boats, river banks, etc., is part of the good practice that follows automatically from this alertness. Dust, powder, skin and hair are obvious external sources of metals, and rigorous care is required to minimize their effects. (ANZECC and ARMCANZ, 2000)

The USGS also provides guidelines regarding the concentrations ranges for which CH/DH sample handling is recommended or required (Table 1). However, regardless of the analyte concentrations present in a study area, USGS notes that CH/DH sample handling is also recommended "when the target analyte could be subject to contamination from field or laboratory procedures at a level that could exceed data-quality requirements" (USGS, 2015). As such, it is of the utmost importance to characterize and understand the entire sample processing blank (*i.e.*, bottle blank + field blank + equipment blank) as early in the process as possible, to determine whether the sample handling techniques being used are appropriate for the study's DQOs. The importance of implementing these practices from the bottle-washing stage through to the analytical stage cannot be overstated. Contamination can occur at any step of the sampling process, from bottle washing, to sample collection, to sample preservation, to sample analysis.

Sampling equipment and equipment cleaning

In addition to using CH/DH sample handling techniques, investigators should also choose appropriate sampling equipment and properly maintain it as a way of preventing or limiting contamination. As mentioned above, clean sampling techniques generally involve the use of sampling equipment that is made of inert, non-metallic components, and when guidance is specific to trace metal analysis, the use of non-metallic sampling equipment is preferred. All Tiers of sample handling techniques outlined in the framework (Figure 1) benefit from the selection of appropriate sampling equipment, but some differences may arise regarding cleaning and equipment care. Many routine sampling devices may contain metal or parts that can potentially leach metals, such as valves, rubber O-rings, clamps, *etc.*, so extra care may be needed to assure that sampling equipment meets the requirements necessary to sample within the study's DQOs. US EPA Region 10 and Puget Sound Water Quality Action Team (1997) suggest using a GO-FLO bottle to minimize potential contamination of deep water samples with surface waters that may be contaminated and also specifies that "Teflon^{TM}-lined Go-Flo bottles are recommended when sampling marine water that will be analyzed for ambient or trace levels of mercury." USGS (2015 Ch. A4) recommends that samplers "Select equipment with components made of fluorocarbon polymer or other relatively inert and uncolored plastics or glass if components will directly contact samples to be analyzed for inorganic constituents. Do not use metal or rubber components for trace-element sampling" (emphasis in original). US EPA Method 1669 also requires that a "metal-free Apparatus" be used for all samples that will be analyzed by inductively coupled plasma mass spectrometry (ICP/MS), hydride atomic absorption (AA), and furnace AA methods (US EPA, 1996a).

Choosing appropriate sampling equipment can be particularly challenging when metals are not the primary analytes of concern in the study or the order of magnitude of metal concentrations expected to be found at the study site is unknown. For example, if the target of a sampling program is assessing NRD at a water body due to oil contamination, the focus of much of the sampling may be to identify hydrocarbons in the water. US EPA (1996a) guidance regarding trace metals states that, "The philosophy behind contamination control is to ensure that any object or substance that contacts the sample is nonmetallic and free from any material that may contain metals of concern." However, sampling for hydrocarbons, as discussed in Kneeland et al. (2020), this issue, requires different sampling equipment and often different bottle types than sampling for trace metal analyses. In order to successfully assess NRD for a water body due to trace metal contamination associated with an oil spill, separate sampling equipment may be required, even if the medium being sampled (*i.e.*, water) is the same. There may be some overlap in appropriate sample bottle material and/or equipment for both hydrocarbons and metal analysis (*e.g.*, equipment made of fluorocarbon polymers is often acceptable for both), but it is important to be aware of the different

sampling requirements of the two analyses when choosing an appropriate sampling protocol for the study. For example, while it is generally accepted that metal or glass sampling equipment and containers are suitable for water sampling of hydrocarbons (Canadian Council of Ministers of the Environment, 2011 p. 27; USGS, 2015, 2010 Ch. A2, p. 16), these materials are generally not acceptable for trace metal sample collection, because they have the potential to contaminate the samples with metals. USGS advises to discontinue use of the US DH-48, US DH-59, US DH-76, US D-74, US D-77, US D-77 Bag, Frame Bag (FB), US P-61, US P-63, and US P-72 samplers for collecting trace-element samples (USGS, 2015). If these samplers are the primary samplers in use for NRD Assessment of other types of analytes, opportunistic sampling for metals may not be possible without a separate set of sampling equipment.

Volume requirements should also be kept in mind during sampling to ensure that the sampling apparatus is capable of collecting sufficient water for the intended analysis. The US EPA Region 10 and Puget Sound Water Quality Action Team (1997) state that, "Regardless of the sampler type, it should have sufficient capacity to supply adequate volume for the tests required." Table 3 describes several different samplers, including their volume, and approaches for both shallow and at-depth sampling. Each is discussed in more detail in the Sampling Devices and Deployment section of this article.

Other water sampling guidance is less specific about the equipment materials and cleaning required, which may be reflective of differences in the DQOs of the guidance. For example, according to US EPA Region 10 and Puget Sound Water Quality Action Team (1997):

> Prior to use, sampling and laboratory equipment should be thoroughly cleaned with a phosphate-free detergent solution, rinsed thoroughly with hot tap water, soaked a minimum of one hour (overnight is recommended) in 20 percent HNO_3, and then rinsed with analyte-free water.
>
> If sampling equipment contains metal components, those parts should be cleaned as stated above, but the acid-soak step should be omitted. If both trace organics and metals analyses are to be performed on the same samples, final rinsing of metal equipment parts with methylene chloride is acceptable.
>
> If trace metals analysis is to be conducted on marine water, the water sampling bottles must not contain metal or rubber parts that could potentially contaminate the water sample. The sampling bottles should be cleaned by first filling them with 20 percent HNO_3 for at least 24 hours, followed by thorough rinsing with metal-free water.

In the above guidance, metal components in sampling equipment are deemed acceptable, and the cleaning protocols do not specify a grade for the acid used to wash the equipment; however, US EPA Method 1638 (US EPA, 1996b) is more specific about the details of bottle, labware, and sampling equipment cleaning, and describes a 10-step process for decontamination that includes the use of detergents, hand-scrubbing, reagent water rinses, acid rinses, extended soaking periods, and wrapping in polyethylene films. This method, while focused on a particular analytical technique, includes details specific to sampling "designed to support water quality monitoring

Table 3. Typical sampling equipment for marine water sampling.

Sampler	Intended Sampling	Sampler Volume	Main Component Materials	Deployment
Go-Flo bottle	Shallow and deep sampling, designed for marine deployment from ships	1.7-100 L	PVC; silicone or viton O-rings	Hydrowire or from a rosette, vertical orientation
Niskin	Shallow and deep sampling, designed for marine deployment from ships	1.2-30 L	PVC; fluorosilicone, silicone, or viton O-rings	Hydrowire or from a rosette, vertical orientation
Niskin-X	Shallow and deep sampling, designed for marine deployment from ships	1.2-12 L	PVC; fluorosilicone, silicone, or viton O-rings	Hydrowire or from a rosette, vertical orientation
Bacon bomb	Designed to sample storage tanks and drums, may be used in other aqueous environments	118-946 mL	Stainless steel, acrylic, or brass; various O-rings	Hand-deployed from a rope or cable, vertical orientation
Kemmerer	General purpose water sampling at depth, water quality sampling, well sampling	0.6-6.2 L	Stainless steel, PVC, acrylic, or Teflon	Hand-deployed from a graduated rope or steel cable, vertical orientation
Van Dorn	Shallow or deep water column sampling, typically estuarine or freshwater	2.2-8.2 L	Acrylic or PVC	Deployed from a rope or cable, horizontal orientation

Note: PVC = Polyvinyl Chloride.

programs authorized under the Clean Water Act" (US EPA, 1996b). The target criteria in this US EPA Method range from 32-0.32 ppb, a concentration range that falls within the Tier II protocols for sample collection (Figure 1, Table 1). If CH/DH sample handling techniques are being used, some sampling equipment cleaning should also be performed. However, the US EPA Region 10 and Puget Sound Water Quality Action Team (1997) guidance quoted above covers both trace-level and generic metal concentration (Supplemental Table 1) and does not specifically state the DQOs. This leaves room for interpretation regarding the applicability of this guidance, which defers to US EPA Method 1669 (US EPA, 1996a) for sample handling *if* trace-level sampling is required. US EPA Method 1669 specifies that equipment should be cleaned with acid, does not mention methylene chloride, and further defers to equipment cleaning procedures specified in whatever analytical US EPA method will be used (see Table 1 in US EPA, 1996a). US EPA Method 1669 also includes the caveat that the method is performance-based and "an alternate sampling procedure or technique may be used, so long as neither samples nor blanks are contaminated when following the alternate procedures" (US EPA, 1996a). Thus, if there is uncertainty regarding the concentration range targeted by the guidance being followed, it is best to follow the guidance back to the analytical method to determine whether more stringent cleaning protocols are required to satisfy a study's DQOs. As one example of sampling equipment cleaning recommendations, if a peristaltic pump is used for sampling, and using Tier II sampling techniques is necessary, the tubing should be cleaned with acid. US EPA (1996a) recommends soaking the tubing in 5-10% HCl solution for 8-24 hours followed by a thorough rinsing and drying step. All tubing should be transported to and from the field double-bagged in plastic.

The choice of an optimum sampling methodology can be impacted by the type of sampling equipment and equipment cleaning methods chosen. Thus, a sampling blank should be tested prior to employing the sampling methodology in the field. For example, as discussed in the excerpt above, both trace organics and metal analyses may be performed on the same sample under some circumstances, as long as any metal equipment parts are rinsed with methylene chloride (US EPA Region 10 and Puget Sound Water Quality Action Team, 1997). To determine whether this sampling approach would meet the study's DQOs, it is advisable to collect a bottle blank before and after the methylene chloride rinse in order to determine whether or not the methylene chloride is a source of metal contamination. If the methylene chloride rinse is necessary for determining trace organics levels, but is a demonstrated source of metal contamination, a separate sample methodology may be required for the metal analyses.

Given the lengths required for the material composition and adequate cleaning of sampling equipment, it is preferable to have dedicated sampling equipment for the sole purpose of sampling for metal analysis. Field decontamination of sampling equipment is considerably more difficult than decontamination in the lab. When possible, it is best to use dedicated or disposable sampling equipment for each field station (*e.g.*, disposable tubing for a peristaltic pump, Teflon bailers). Some types of sampling equipment, such as a Kemmerer sampler, may be particularly difficult to decontaminate in the field (ASTM, 2016a). US EPA recommends that, when a field apparatus must be reused in the field, the apparatus should be rinsed with dilute nitric acid, then rinsed with 1 L of reagent water, a field blank should be collected, and then the sampler should be rinsed again with "copious amounts of the ambient water sample" prior to sample collection (US EPA, 1996a). If having two sets of sampling equipment on-hand during sampling is not possible, it may be advisable to clean the sampling equipment being used for collecting metal analyses samples in between sampling campaigns, or to clean this sampling equipment prior to field analysis and carry out all metal sampling before sampling for other analytes.

In addition, sample processing in the field should take place in as "clean" an environment as possible. Available sampling guidance often recommends that sample processing and subsampling occur in isolated, Class 100, or other "clean" environments. In the field, sample processing (*e.g.*, filtering, acidification) may be required on a shorter timeframe than can be achieved with immediate transport back to the lab. There are several ways in which clean sample processing can be carried out in the field. Field-portable glove bags, purchased or "constructed with a nonmetallic (PVC pipe or other suitable material) frame and a frame cover made of an inexpensive, disposable, nonmetallic material (*e.g.*, a thin-walled polyethylene bag)" can be employed to minimize contamination when performing sample processing in the field, such as filtering samples for dissolved metals (US EPA, 1996a). Larger Class 100 clean environments for processing trace-level samples can also be constructed in the field, on ships or at field sites. Such environments can also be constructed with non-metallic material, such as polyethylene sheeting,

and outfitted with high-efficiency particulate air (HEPA) filters to keep air particulate concentrations low and maintain a positive pressure environment, and allow them to be used as a temporary walk-in laboratory (Cutter and Bruland, 2012).

Finally, as alluded to above, some analytes can be considered above their criteria or screening levels at very low concentrations (*e.g.*, Hg). Sampling practices for Hg are among the most stringent. As noted in US EPA Method 1669, an "unlined, long-sleeved wind suit consisting of pants and jacket and constructed of nylon or other synthetic fiber is worn when sampling for mercury to prevent mercury adsorbed onto cotton or other clothing materials from contaminating samples" (US EPA, 1996a). Any ancillary sampling equipment or other supplies required when sampling for certain analytes should be considered during the planning stage of the sampling program. Their necessity, as discussed earlier, is dependent upon the study DQOs (Figure 1).

Sample container selection

Another crucial decision to make when assembling a trace metal sampling plan is to choose appropriate bottles for storing samples. To simplify this decision, many guidance documents recommend that sample bottles (including appropriate preservative, when applicable) should be provided by the laboratory that will be analyzing the samples (*e.g.*, ADEQ, 2015). This allows the experts in the analytical procedure to take responsibility for ensuring that the bottles are made of an acceptable material and are cleaned appropriately for the analyses to be performed.

If this is not possible and the sampling plan must include bottle purchase and preparation, several further decisions may be required. Again, these decisions are performance-based. To ensure a successful and efficient sampling campaign, it is advisable to perform bottle blank analyses to determine whether the sample bottles and cleaning procedures employed are appropriate for the analyte concentrations expected to be found in the field and the criteria that need to be met. Conducting pre-field campaign performance assessments can save significant time, money, and resources. For example, if a bottle blank performed on a new sample bottle produces analyte concentrations that are adequately below the minimum detection limit, screening criteria, or background level required to achieve the DQOs, then acid washing or more stringent bottle preparations may not be needed, and considerable time and resources can be saved by eliminating these steps. On the other hand, if bottle blank concentrations are too high or inconsistent, the sample bottles may need more rigorous cleaning or preparation prior to sampling. Identifying this need before a field sampling campaign begins can significantly reduce the possibility that the sampling campaign would need to be repeated to collect representative and defensible samples. The key decisions regarding appropriate sample bottle collection and preparation are discussed below.

First, while in some circumstances, re-using sample bottles is permitted after thorough washing (Government of Western Australia, 2009), most sampling guidance specifies or assumes that new sample bottles will be used, and used only once. For bottles that are re-used, rigorous cleaning procedures are often recommended, requiring the sampler to at least "rinse with detergent (De-Con 90 is recommended), then very thoroughly wash and rinse with deionised or distilled water…. Other washing solvents include dilute hydrochloric acid (HCl) (0.1 moles/L HCl)" (Government of Western Australia, 2009). If possible, new bottles should always be used for sampling.

Second, assuming new bottles are used, the material of the sample bottles is critical, particularly for Tier II and Tier III sample handling techniques (Figure 1). For most routine trace metal analytes, some form of plastic, usually high- or low-density polyethylene (HDPE/LDPE), bottle is sufficient. However, bottle requirements may be different for some metals, and it is important to understand whether or not all the analytes of interest can be successfully measured from the same type of bottle. Bottle requirements vary by agency/organization, with bottle material ranging from glass to Teflon, and bottle sizes ranging from 60 mL to 1 L. Some water sampling guidance offers the sampler a choice of bottle material (*e.g.*, glass or plastic) for metal analyses, while the State of Washington specifies that only Teflon bottles be used for metal analyses (WADOE, 2010). Guidance focused on the reliable detection of low ppb and ppt concentrations are more specific about the bottle material required. There is general consensus that while using glass bottles is sometimes acceptable for sampling trace metals, Teflon is the only appropriate *plastic* material for Hg sampling equipment, and using polyethylene, polycarbonate, and other non-fluoropolymer equipment for this purpose is explicitly discouraged (US EPA, 1996a; US EPA Region 10 and Puget Sound Water Quality Action Team, 1997; US EPA Region 6, 2000; ANZECC and ARMCANZ, 2000; OWRB, 2003; California State Water Resources Board, 2007; LDEQ, 2008; NJDEP, 2011; Canadian Council of Ministers of the Environment, 2011; TCEQ, 2012; FLDEP, 2014a).

To sample water for the analysis of silver (Ag), the Oklahoma Water Resources Board (OWRB, 2003) specifies that glass amber bottles be used instead of plastic, and to sample for selenium, the Australian and New Zealand Environment and Conservation Council (ANZECC) and Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ) specifies that plastic bottles should be used, excluding polycarbonate and some types of polyethylene bottles (ANZECC and ARMCANZ, 2000).

Requirements to use specific sample bottle types are often driven by differences in the chemistry of the different metals. Some metals are redox-active (*e.g.*, Cr), others are light-sensitive (*e.g.*, Ag), volatile (*e.g.*, Hg), and/or easily complexed to organic matter (*e.g.*, Cu). For the majority of metals, however, polyethylene bottles are sufficient, and some form of preservation (most commonly acidification) is sufficient to keep the metals dissolved in solution and preserved in a form that will allow for the determination of representative analytical results. While the use of fluoropolymer sample bottles is generally considered the best option for most metal analytes, the potential benefit of their use over polyethylene (with the exception of Hg, for which fluoropolymer bottles are required) is frequently outweighed by the much higher cost of fluoropolymer bottles. Clear or colorless plastic materials should be used rather than colored plastics, because dyes may contain metals or other sources of contamination. It is also good practice to assure that the entirety of a sample bottle is made of an acceptable material: *e.g.*, " … container lids should be checked for liners that may cause contamination or adsorb particular analytes" (ANZECC and ARMCANZ, 2000). An example of a common liner that may cause contamination would be a rubber O-ring seated inside the bottle cap to improve the seal, which has the potential to contaminate the sample with Zn.

Third, there is general consensus among water sampling guidance that some level of sample bottle cleaning is required prior to sample collection, even for new bottles (note that this decision can likely be avoided if the laboratory is providing the sample bottles). Commonly, sampling guidance specifies that acid should be used for bottle washing, in order to leach any metals that may be present in the bottle material itself. This, however, is often where the consensus ends. Guidance details vary on the acid type (generally HCl, nitric acid [HNO_3], aqua regia), concentration (dilute, 10%, 50%, *etc.*), and purity (unspecified, reagent grade, ultrapure, *etc.*), as well as the duration (rinse, overnight soak, multiple-day soak) of each of the cleaning steps. Some cleaning procedures call for heat at different steps in order to speed up the cleaning process (VADEQ, 1996). For example, the cleaning recommendations provided in US EPA Method 1669 are as follows:

> Before samples are collected, all sampling equipment and sample containers are cleaned in a laboratory or cleaning facility using detergent, mineral acids, and reagent water as described in the methods referenced in Table 1. The laboratory or cleaning facility is responsible for generating an acceptable equipment blank to demonstrate that the sampling equipment and containers are free from trace metals contamination before they are shipped to the field sampling team … After cleaning, sample containers are filled with weak acid solution, individually double-bagged, and shipped to the sampling site. All sampling equipment is also bagged for storage or shipment … EPA has found that, in some cases, it may be possible to empty the weak acid solution from the bottle immediately prior to transport to the field site. In this case, the bottle should be refilled with reagent water … (US EPA, 1996a)

The purpose of keeping cleaned sample containers filled with weak acid solution is to neutralize potential adsorptive sites with hydrogen ions and minimize negative contamination from adsorptive losses. It is best practice, whether the bottles are new, cleaned, or supplied by the analytical laboratory, to collect and test a bottle blank to ensure that whatever cleaning procedure is adopted will adequately meet the study's DQOs. This is yet another example of the many ways in which best sampling practices are explicitly performance-based.

Finally, handling and transporting the bottles themselves is important to consider with respect to contamination prevention. The Canadian Council of Ministers of the Environment (2011) specifies that "Sample bottles should be kept in a clean environment, away from dust, dirt, fumes, and grime. As well, bottles must be capped at all times and stored in clean shipping containers … " Sampling guidance often recommends that samples suspected to have very different (order of magnitude) metal concentrations be stored and shipped separately, to avoid contamination of the lower-concentration samples (Canadian Council of Ministers of the Environment, 2011).

Sampling devices and deployment

Guidance for water sampling for trace metals is similar to that for other water sampling procedures, with the primary distinction being the requirements for using CH/DH techniques and specific equipment

materials. These requirements add extra layers of complexity and precaution to the trace metal sampling process, which should be performed by experienced and trained personnel. Most of the deviations from general sampling procedures in trace metal sampling guidance are aimed specifically at reducing contamination. For example, when multiple sites will be sampled during a single field excursion, sites with the least expected contamination or lowest concentrations of analytes should be sampled first, when possible (*e.g.*, USGS, 2015). If sampling is being done from a ship, the "vessel should ideally be positioned downwind or down current of the sampling device," and "care should be taken to avoid visible surface slicks and the vessel's exhaust" (US EPA and US ACE, 1995). Surface slicks and films can be a potential source of contamination, but US EPA Region 10 and Puget Sound Water Quality Action Team (1997) warn that "not all surface microlayer contamination will be in the form of visible slicks."

Water sampling devices can be divided into "isokinetic" and non-isokinetic" types. Isokinetic sampling devices can be used to obtain representative samples of flowing water sources (*e.g.*, tidal zones, estuaries) and are designed such that the flow of water entering the sampling device is the same as that of the current. Isokinetic sampling is believed to increase the representativeness of samples when particulates are a concern (USGS, 2015). However, a summary of such sampling methods is beyond the scope of this review. For further information on methods of sampling flowing water sources, see Chapter 4 of the USGS *National Field Manual for the Collection of Water-Quality Data* (USGS, 2015). All the samplers described below are considered to be non-isokinetic sampling devices.

Sampling in the marine environment can be performed at the surface during sampling campaigns from small boats, wading, or the shore. Sampling can also be performed at the surface and to shallow and deep depths further from the shore. Offshore operations may involve more complex sampling and the use of dedicated machinery on ships. The sections below describe the recommended sampling devices and protocols to use near the surface and at depth, as well as ancillary data that may be useful to collect at the time of sampling.

Sampling at or near the surface

"Surface water" in this review refers to water at or near the surface of a water body and not to the larger-scale use of the term (*i.e.*, "All bodies of water on the surface of the earth," as defined in *McGraw-Hill Dictionary of Scientific and Technical Terms, Third Edition*; Parker, 1984).

There is no specific consensus on what depth a sample should be collected from to constitute a surface or subsurface sample. For example, State of Texas water sampling guidance (TCEQ, 2012) refers to "near surface" samples as those collected at a depth of 0.3 meters, while State of California guidance (California State Water Resources Board, 2007) describes "subsurface" samples as those collected at a depth of 0.1 meters. Most of the water sampling guidance documents reviewed state that a surface water sample should be collected from just below the surface of the water body but do not specify how deep the sampling device should be submerged during sample collection. In the open ocean, it may also be difficult to determine the depth of a shallow surface sample if seas are high. When possible, surface sampling depths should be "consistent and documented" throughout the sample collection process (USDA, 2001). It should be noted that "surface" and "subsurface" do not refer solely to the surface micro-layer of the water body (*i.e.*, "the top layer of surface waters, approximately 50 microns thick"), which can be enriched in certain metals and organometallic compounds and is sampled infrequently and only with specialized sampling devices, such as a "drum roller" sampler (Canadian Council of Ministers of the Environment, 2011). Discussion of sampling the surface micro-layer is beyond the scope of this review.

While surface waters can be sampled by many means, the simplest way to sample surface water is by collecting "grab samples." Surface water grab samples are shallow samples that may be collected by hand (*i.e.*, using the sample container itself as the sampling device), with a sample jar or intermediate vessel attached to a pole, or by using a bucket or bailer (Figure 2A-D; US EPA, 1996a; TCEQ, 2012). Grab samples are not limited to shore-based sampling and may also be collected from the side of a small boat. Sampling surface water by submerging the sample container itself should be done carefully, so that a representative sample can be obtained and contamination can be avoided. For instance, USGS (2015) notes that, "when sampling with a hand-held bottle … stand downstream of the bottle while it is being filled … Care must be taken to avoid collecting particulates that are resuspended as the result of wading or bumping the sampler on the streambed." This guidance is specific to streams, but the same principles apply to the marine environment: one should avoid sampling

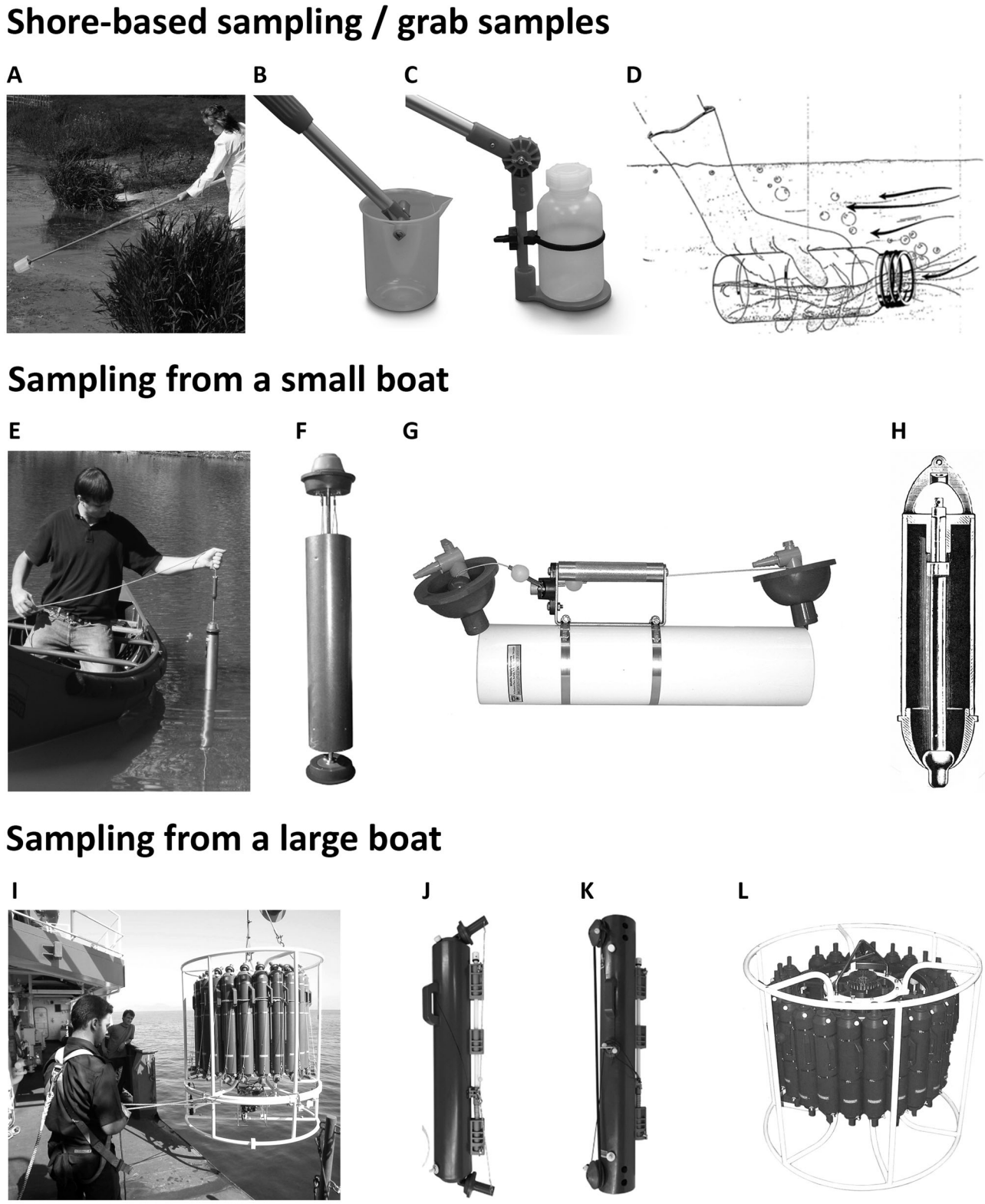

Figure 2. Examples of water sampling equipment and deployment. (A) Deployment of a pole sampler while sampling from shore (pole samplers may also be deployed from small boats or ships) (Source: Sampling Systems Ltd., 2017). Different attachments may be used with a pole sampler, such as (B) a swinging beaker sampler (Source: Sampling Systems Ltd., 2017), or (C) a sample bottle (Source: Sampling Systems Ltd., 2017). (D) Grab samples may also be collected from shore or when sampling from a small boat by carefully submersing the bottle beneath the surface and facing the direction of flow (Source: US EPA, 1982). (E) Example of sampling from the side of a small boat (Source: Wildco, 2016a). Sampling at shallow or intermediate depths can be achieved with several samplers, including: (F) Kemmerer-type samplers (Source: Wildco, 2016a), (G) Van Dorn-type samplers (Source: Wildco, 2016b), and (H) Bacon Bomb-type samplers (Source: Gammon Technical Products, Inc. 2017). (I) For deeper and often larger-volume sampling from ships (Source: Moll, 2017), samplers like the (J) Niskin bottle (Source: General Oceanics, Inc., 2016a) or (K) GO-FLO bottle (Source: General Oceanics, Inc., 2016b) can be deployed either individually or (L) around a rosette (Source: Sea-Bird Scientific., 2017).

within the exhaust wake of the ship, and with the sampling container facing upcurrent. When collecting grab samples from a boat, the Canadian Council of Ministers of the Environment (2011) instructs the sampler to "Plunge the bottle, neck downward, below the surface to a depth of about 20 cm. Immediately turn the bottle until the neck points slightly upwards with the mouth directed into the current.. Hold the bottle facing upstream at arm's length while it fills." If grab samples are being collected near a surface slick, the sample technician should "Immerse [the] sample container with the opening pointing directly down to maintain a volume of air in the container, thereby avoiding the collection of any surface films" and potential sample contamination (Australia Northern Territory Government, 2009). If an intermediate sampling device is used to collect surface/subsurface water samples, the sampled water should be decanted immediately into the sample bottles appropriate for the desired analytes (*i.e.*, subsampled). US EPA Region 10 and Puget Sound Water Quality Action Team (1997) recommend that this be done "as soon as possible (*i.e.*, within 15 minutes), as appreciable delay may result in unrepresentative subsamples."

Water at or near the surface may also be collected using a peristaltic pump to pump water directly into sample bottles (Canadian Council of Ministers of the Environment, 2011; USGS, 2015). This sampling option allows for both collecting larger-volume samples and collecting samples directly into sample containers to avoid contamination. The most common pump tubing types are flexible tubing used with the peristaltic pump itself (composed of styrene ethylene butylene styrene [SEBS] resin, with a 3/8-inch or 1/4-inch inner diameter) and stiffer tubing used for the sampler used with the pump (composed of fluoropolymer, also with a 3/8-inch or 1/4-inch inner diameter) (US EPA, 1996a). If sample filtration is required, filters can also be affixed directly to the sampling tubing train. However, while this type of system may help prevent contamination, the International Organization for Standardization (ISO) notes that it is unsuitable for collecting volatile compounds (*e.g.*, Hg) (ISO, 1992). Additionally, the USGS notes that, "A disadvantage of collecting a sample by pumping is that if a thin stratum of water is being sampled, water can move radially from unknown depths and distances into the pump" (USGS, 2015). When sampling for trace metals in water, some guidance specifies that direct sampling into a sample container is preferable, because "This procedure is the simplest and provides the least potential for contamination because it requires the least amount of equipment and handling" (US EPA, 1996a).

Sampling at depth (profiling)

Water column sampling is done to obtain representative samples from discrete depths at a pre-determined sampling point or from a number of discrete sampling points to create a vertical or horizontal profile of metal concentrations (USGS, 2015). Vertical profiles may also sometimes include a near-surface sample.

A variety of sampling devices can be employed to collect discrete water column samples for trace metal analysis. The types of devices range from off-the-shelf intermediate sampling devices, to devices that sample directly into a sample container, to peristaltic pump systems. Although commercially available devices are widely available, some water sampling guidance allows for custom-made sampling devices to be used as well (US EPA, 1982). Some sampling equipment may be more well- or ill-equipped to achieve the DQOs of a study than others, so it is important to compare the features of different samplers and understand how they might help or hinder the sampling effort's objectives.

Sampling devices like the Niskin (Figure 2J), Kemmerer (Figure 2F), and Van Dorn (Figure 2G) water samplers are deployed open until they reach the designated depth, then a messenger is used to close the sampler. Thus, the interior surfaces of these devices are exposed to the entire water column as they pass through it to reach the sampling location. This can be a disadvantage when the levels of trace metals expected to be encountered are low and/or if the sampling device will pass through a surface layer that may be contaminated with the metal(s) of concern (ASTM, 2016b). Sampling devices like the liquid grab sampler (not pictured), GO-FLO bottle (Figure 2K), point sampling bailer (not pictured), Bacon Bomb (not pictured), and lidded sludge/water sampler (not pictured) were designed to remain closed during initial deployment through the surface layer. These devices open either below the surface (*e.g.*, GO-FLO bottles are pressure-tripped to open at a depth of 10 m; US EPA Region 10 and Puget Sound Water Quality Action Team, 1997) or once the designated sample depth has been reached. The sampling device is then tripped to close, ensuring that its interior does not encounter a potentially contaminated surface layer or contaminated waters located above the chosen sampling depth (ASTM, 2016b; US EPA Region 10 and Puget Sound Water Quality Action Team, 1997). It should be noted that while some authoritative guidance discusses the

use of Bacon Bomb samplers (US EPA, 1994), the tripping mechanism for these samplers is triggered when it hits bottom, and these samplers are more frequently used to collect water from within drums or tanks. Hitting bottom is not recommended or appropriate for most marine sampling, because hitting the bottom stirs up sediment that can contaminate the sample and marine samples are not often collected at the bottom of the sea floor.

As with surface/subsurface sampling, using sampling techniques that collect samples directly into the sample bottle or employ intermediate samplers for collecting samples have both advantages and disadvantages. Liquid grab samplers, weighted sampling bottles, lidded sludge/water samplers, and simple extension poles allow for the collection of samples directly into the sample container, while most off-the-shelf sampling devices are designed as intermediate samplers, which require an additional step of subsampling from the sampler into a sample container appropriate for the analyte of interest. A sampling device that collects a sample directly into the sample container "eliminates sample contact with conventional samplers… thereby reducing the risk of extraneous contamination" (US EPA, 1996a). While such samplers may be useful for small sampling endeavors in areas where reducing sample contamination is already a challenge, their ability to collect sufficient sample volume may be limited, particularly if several ancillary samples and/or discrete samples for different trace metal analyses need to be collected from the same location. Furthermore, if the ability to deploy several samplers during one cast is desired, using an intermediate sampler (many of which have this capability) would be necessary.

To more efficiently collect many samples at depth during one cast, US EPA Region 10 and Puget Sound Water Quality Action Team (1997) suggest that "Multiple water samplers can be attached sequentially to a vertical hydrowire for sampling at multiple depths on a single cast, or they can be mounted on a rosette frame (often in conjunction with an *in situ* sensor array) which allows for collection of replicate samples at the same depth." The rosette or hydrowire with the attached samplers is then lowered through the water column (US EPA Region 10 and Puget Sound Water Quality Action Team, 1997). When the sampler is at a desired sampling depth, a messenger or remotely operated device is used to close each individual sampler before the rosette or hydrowire is brought back to the surface. If a rosette is used, samples may be collected sequentially while the rosette is brought up through the water column, thus effectively sampling an entire depth profile of water in one cast. Rosettes are sometimes capable of holding as many as 36 sampling bottles. US EPA Region 10 and Puget Sound Water Quality Action Team (1997) advise that, once at depth, the sampler "should be allowed to equilibrate to ambient conditions for approximately 1 minute before it is closed," and that if a profile is the desired sampling product, "it is advisable that [the samples] be collected from a single cast" in order to ensure good representation of the targeted water parcel. This guidance further stresses that a representative sample profile of the entire water column cannot be accomplished with multiple casts, because ambient conditions change rapidly near the water surface. However, if a sampler fails at depth, it is deemed acceptable for a second cast to be performed, because conditions change more slowly deeper in the water column (US EPA Region 10 and Puget Sound Water Quality Action Team, 1997).

Once a discrete-depth sampler has been retrieved and brought back onboard the ship, sample bottles are filled from a drain valve near the bottom of the sampler. Before dispensing a water sample from the sample bottle, "each bottle should be checked immediately for leakage of sample water around the seals… If the sample has been compromised, the cast should be repeated" (US EPA Region 10 and Puget Sound Water Quality Action Team, 1997). During the subsampling stage, it is also important to determine whether the sample bottles should be field-rinsed prior to sample collection from the sampling device. The US Forest Service recommends that immediately prior to sampling, "the bottle and cap [be] rinsed three times with sample" (USDA, 2001). Field-rinsing is mandated by the ISO Standards for water sampling (ISO, 1992, p. 6), but can also be explicitly prohibited if a laboratory has provided sample bottles with a premeasured amount of preservative (Government of Western Australia 2009). In general, field-rinsing is considered good practice if sample bottles are not shipped to the field with preservative. As noted above, when employing CH/DH techniques, subsampling may be carried out in a clean environment, such as a glove box or a clean laboratory under positive pressure. The decision about whether or not to subsample in a clean environment should align with the DQOs of the study. To collect samples for determining compliance with US EPA NRALWQ Criteria (Table 1), subsampling in a clean environment is strongly recommended.

Yet another option for sampling at depth is using a peristaltic and/or submersible pumping system that can be lowered to predetermined depths (USGS, 2011; US EPA Region 10 and Puget Sound Water Quality

Action Team, 1997; US EPA, 1996a). However, the length of the pump tubing may be a limiting factor for some sampling plans, and, as reviewed in the previous section, some water sampling guidance notes certain disadvantages to using a pump and gives preference to other sampling devices.

Ancillary data

In addition to the analyte(s) of interest, other sampling information and metadata may be informative or even necessary for successfully interpreting analytical results. If possible, it is good practice to document site conditions and collect ancillary data prior to collecting water samples. Most water sampling guidance recommends measuring temperature, pH, dissolved oxygen, conductivity (specific conductance), and water level or depth (*e.g.*, ADEQ, 2012; USDA, 2001; Australia Northern Territory Government, 2009). *In situ* detection of many of these analytes can be achieved by deploying a sensor array along with the sampling device, most often *via* a hydrographic wire or conducting cable. The most basic arrays, referred to as CTDs (which stands for conductivity, temperature, and depth), measure conductivity (which can be converted to salinity and used to calculate density), temperature, and water pressure (which is used to determine depth) (US EPA Region 10 and Puget Sound Water Quality Action Team, 1997). US EPA Region 10 and Puget Sound Water Quality Action Team (1997) note that, "Additional sensors can be included to measure other water column variables such as dissolved oxygen, pH, irradiance, turbidity, oxidation-reduction potential and chlorophyll *a*." These data may help with interpreting the potential behavior and predicted cycling of metals, as well as with determining their potential impact on the surrounding ecology. Ancillary data measured *via* a sensor often need to be calibrated using shipboard analyses (*e.g.*, winkler titration for oxygen [O_2] analyses). For all calibration measurements, calibration must be performed on a separate aliquot and not on the sample collected for trace metal analysis.

Analyses of some metal species require additional ancillary data collection and analysis. It is important to understand whether the criteria used for NRD Assessment account for variables such as metal speciation, pH, or organic matter content. The parameters accounted for in the criteria and those collected should be coherent. Otherwise, the criteria may be overly stringent or underprotective, depending on the parameter and how the criteria and analyte are compared. For example, Cu has a very low US EPA NRALWQ Criteria value (Table 1), but this is reflective of the understanding that Cu bioavailability and toxicity is primarily related to the free cupric ion, Cu^{2+} (US EPA, 2016a). Cu is a redox-active element with several possible oxidation states (Cu^0, Cu^{1+}, Cu^{2+}, Cu^{3+}) and forms complexes to organic matter (US EPA, 2016a). In oxic, neutral waters, Cu^{2+} is generally a minor fraction of the total Cu present, while the majority is present in non-bioavailable forms bound to organic or inorganic compounds in solution either as complexes or precipitates. In a routine total metals sample, preservation of the sample by acidification will cause dissociation of Cu from ligand complexes, dissolution of Cu precipitates, and transformation of any dissolved Cu present into Cu^{2+}. The preservation process fundamentally changes the potential bioavailability and toxicity of the metal. Thus, comparing unfiltered total Cu to the NRALWQ Criteria would overestimate Cu toxicity. In recognition of this finding, US EPA provides guidance, based on the BLM, for estimating Cu^{2+} concentration from total dissolved Cu concentration. In order to perform this calculation, other inputs are needed to satisfy the model. For marine waters, this includes temperature, pH, dissolved organic carbon, and salinity. A full discussion of the ancillary needs for a Cu BLM is beyond the scope of this review, but a more in depth discussion can be found by consulting the US EPA BLM guidance (US EPA, 2016a).

Sample processing (filtration and preservation)

Filtration

Some water samples require filtration after collection. Water samples may be analyzed unfiltered to determine total metal concentrations in the sample, but filtered samples are required for determining dissolved metal concentrations. Some project DQOs may require both total and dissolved metal concentrations be determined. As alluded to above, if a pump is used to collect filtered samples or transfer them from the sampling device to the sample container, a capsule filter can be added to the sampling train (US EPA, 1996a; TCEQ, 2012), which eliminates a separate filtration step and minimizes potential sample contamination. Filtration can also be performed offline. Samples will require offline filtration if they were collected *via* a non-pump sampling device, such as a GO-FLO bottle or a handheld sample container. If this sampling approach is taken, State of Texas water sampling guidance describes how to minimize

contamination when filtering in the field by performing filtering within clean plastic bags (TCEQ, 2012). It is recommended that "Samples … be filtered as soon as possible after sample collection, preferably on site" (Government of Western Australia, 2009). This will limit any changes to metal concentrations due to microbial activity, oxidation, or adsorption to the walls of the sample bottle.

When determining dissolved metal concentrations is required, US EPA (1996a) recommends that "samples … be filtered through a 0.45 μm capsule filter at the field site" to separate potentially metal-bearing particulates from the dissolved metals (see also ADEQ, 2012; ORDEQ, 2013). It is important to keep in mind that "dissolved" is ultimately an operational definition that is determined by the filter pore size. As a result, maintaining filtration consistency to ensure comparability throughout a study and against the criteria of interest is critical, as is keeping this in mind if one is comparing sampling results to past results that may have had different filtration criteria. The US EPA NRALWQ Criteria are based on dissolved metal concentrations; however, nowhere in 40 CFR 131 is there a definition of "dissolved." This lack of specificity leaves room for potential unanticipated discrepancies, or inappropriate comparison between datasets. US EPA (1996a) specifies that the filter used for sample filtration should have a diameter of at least 15 mm (and recommends using Gelman Supor 12175 filters or their equivalent), with a 0.45 µm pore size. For consistency with US EPA's recommendations, we recommend using a 0.45 µm filter.

Similar to the considerations for sample bottle material, the materials some filters are composed of are not suitable for the analyses of some metals, and thus it is important to confirm that the filters used are appropriate for the analytes of interest. For example, US EPA (1996a) specifies that 0.45 µm Gelman Supor filters should be used for filtering samples being analyzed for most metals, but specifically stipulates that different filters need to be used for filtering samples that will be analyzed for CrIII (0.4 µm pore size, 47 mm filter diameter, polycarbonate Nuclepore or equivalent). As with sampling equipment, filters and filtration equipment often require cleaning. Filter-cleaning protocols and filter blank testing should also align with the study's DQOs. These can include a soak or rinse with dilute acid to clean the filters prior to use, and a subsequent collection of a filter blank. As with other guidance variability, the duration, type, and purity of acid may vary. Additionally, the filters should not touch metal. For instance, if using forceps is required for seating filter papers in the filter cartridges, the forceps should be plastic, not metal (and, for samples that will be analyzed for CrIII, US EPA [1996a] specifies that fluoropolymer forceps should be used during filter cleaning). Forceps and any portion of the filtration equipment that will come in contact with the water or the filter may also require cleaning similar to those procedures required for sample bottles or sampling devices. Lastly, sample filtration, like other steps in the sampling process, benefits from sampler experience and training. For best results, change filters between samples, do not overload filters, and, whenever flow rate decreases, do not over pressurize or allow filters to run dry.

Sample preservation

Samples collected for trace metal analyses require preservation. While some guidance documents do not provide specifics regarding sample preservation procedures and defer to project documents such as the Quality Assurance Project Plans (QAPPs) or other study-specific sampling and analytical plans (*e.g.*, ANZECC and ARMCANZ, 2000), most authoritative guidance documents offer specifics (*e.g.*, US EPA Region 6, 2000). The most common preservation technique for routine trace metal analyses requires adjusting the pH of the sample to <2 using acid, to keep metals in solution and prevent metals from adhering to container surfaces (for more details, see Supplemental Table 1). This acidification process is intended to prevent negative contamination from adsorptive losses to active sites on the insides of the sample bottle by neutralizing any potential adsorptive sites with hydrogen ions. The amount of acid needed to acidify a sample may vary with water volume and chemistry; testing a split sample can be used to determine the amount of acid required. If a separate sample cannot be test-acidified, a sacrificial aliquot poured from the sample bottle should be used. pH sensors or any other probe should never be placed inside the actual sample bottle (OWRB, 2003; US EPA Region 1, 2003), as doing so poses a high risk of sample contamination.

Nitric acid is specified for most preservation protocols, although some guidance states that HCl can also be used, and other guidance specifically requires that HCl be used if the sample will be analyzed for Hg (California State Water Resources Board, 2007; NJDEP, 2011) or Hg species (US EPA Region 10 and Puget Sound Water Quality Action Team, 1997). The Western Australia Department of Water guidance's (Government of Western Australia, 2009) requirements for samples collected for dissolved Hg analysis are unique, in that it specifies that, "After filtration

add 2 mL of a 20% solution of potassium dichromate in approximately 4 M nitric acid per litre of sample water." In this case, the sample treatment and preservative is specific to an alternative analytical method for measuring Hg. Guidance documents that are concerned with the detection of low ppb and ppt metal concentrations specify that ultrapure nitric acid be used for preservation to ensure that the acids used for preservation do not contaminate the samples (US EPA, 1996a; California State Water Resources Board, 2007; LDEQ, 2008; TCEQ, 2012). Other guidance that is less specific or tailored to detection of high metal concentrations either state only that the acid used for preservation be reagent grade or do not specify the grade of the preservation acid at all (Applied Marine Sciences, 2013; NJDEP, 2011). Some guidance documents also require that, following preservation, samples be put on ice or refrigerated (US EPA Region 10 and Puget Sound Water Quality Action Team, 1997; LDEQ, 2008). Again, the differences among these preferences can often be traced to the study's DQOs, and thus, the necessary Tier of sample handling technique required. For Tier II and III, ultrapure acids are often recommended or required.

It is often assumed that preservation will occur in the field soon after sampling to prevent any adsorptive losses to the sample container surfaces. However, some water sampling guidance documents recognize the complications associated with shipping nitric acid to the field and the return shipment of acidified samples. These guidance documents allow for the acid preservative to be added once the samples have been received by the laboratory (TCEQ, 2012; NJDEP, 2011; California State Water Resources Board, 2007). US EPA offers additional steps when a sample is not preserved in the field:

> Store the preserved sample for a minimum of 48 hours at 0-4 °C to allow the acid to completely dissolve the metal(s) adsorbed on the container walls. The sample pH should be verified as <2 immediately before withdrawing an aliquot for processing or direct analysis. If, for some reason such as high alkalinity, the sample pH is verified to be >2, more acid must be added and the sample held for sixteen hours until verified to be pH <2. (US EPA, 1996c)

While samples slated to undergo routine trace metal analyses are generally preserved with nitric acid, "Preservation choices will vary depending on the parameter to be measured" (ANZECC and ARMCANZ, 2000). So, while general sample preservation practices may call for acidification with nitric acid, it is also important to understand the preservation requirements for all of the metals of interest. If different preservation protocols are required for different metals that are a part of the study's DQOs, separate samples need to be collected, and this should be accounted for when determining the volume requirements for the project plan.

Maximum holding times are almost uniformly 6 months for preserved samples of metals other than Hg and 28 days for preserved samples designated for Hg analysis. US EPA and US ACE guidance from 1995 specified a maximum holding time 14 days for samples being analyzed for Hg (US EPA and US ACE, 1995), but the majority of other US EPA guidance specifies the 28 day holding time (US EPA, 1982, 1986; US EPA Region 6, 2000; US EPA Region 10 and Puget Sound Water Quality Action Team, 1997). However, the Western Australia Department of Water stipulates that samples to be analyzed for Hg to be held for 6 months maximum if frozen and 1 month maximum if refrigerated (at 1-4 °C) and acidified to pH <2 (Government of Western Australia, 2009).

Quality control samples

While several water sampling guidance documents define the nature of QC samples and discuss their importance, many do not provide specific guidance regarding the number and type of QC samples that are considered "sufficient." As with preservation techniques, many guidance documents direct the sampler to more specific guidance, such as QAPPs or sampling plans for the specific sampling campaign, for further details and recommend developing QC sampling plans specific to the study and its DQOs. For further discussion and guidance on the development of an acceptable QC sampling plan for a study, please refer to Wait et al. (2020), in this issue. Table 2 of this article can be used as general guidance on what types of blanks can be taken to quantify contamination associated with each component of the sample collection process. At the very least, bottle blanks (to assess contamination from the sample storage bottles), field/trip blanks and equipment blanks (to assess contamination from the sampling equipment and the rigor of a sample handling procedure), and process blanks (*e.g.*, filter blanks and acid blanks to assess sources of contamination from sample processing) should be considered critical to the success of a trace metal sampling program.

In Appendix B of its "Guidance Document for the Development of Site-Specific Water Quality Criteria for Metals" (OWRB, 2003), the State of Oklahoma identifies seven aspects of sampling that should be

used to ensure the representativeness of samples collected for trace metal analysis:

1. Clean techniques for collection, handling, storage, preparation, and analysis (to avoid contamination).
2. Analytical methods with appropriate detection limits.
3. Analytical methods that avoid interference.
4. Blanks (to assess contamination).
5. Replicates (to ensure precision).
6. Certified standards (for confirmation and calibration).
7. Matrix spikes (to assess interference and contamination).

Sufficient QC samples should be collected to address each of the above aspects.

Summary and conclusions

Although water sampling guidance varies, selecting an appropriate sampling protocol for trace metal NRD Assessment in the marine environment is possible if the study's DQOs are well defined. The DQOs should be used to: (1) select a suite of metal samples that will generate appropriate and meaningful data for the assessment, (2) determine the detection level requirements, and (3) select the analytical techniques to be used. As described in our proposed framework for choosing a sampling approach for trace metal analysis in marine waters (Figure 1), the answers to these questions should be used to decide what type of samples are required (total metals, dissolved metals, specialized metal samples), the level of cleanliness required for sample collection (*i.e.*, Tier I, II, or III sampling techniques; Figure 1), and the type of sample preservation required to ensure that the sampling process will produce representative samples that can be analyzed using the selected analytical methods. While making these decisions, it is important to keep in mind that concentrations that cause NRD are metal-dependent and also span several orders of magnitude. The analytical techniques available may have detection limits much lower or higher than the blanks associated with typical sampling practices.

Using the study's DQOs as a guide, our framework for choosing a sampling approach can help investigators make informed decisions about the sampling and sample processing procedures needed to successfully collect reliable and defensible trace metal water samples. DQOs are assumed to include the type(s) of metal samples to be collected, the criteria against which the samples will be evaluated, and the analytical techniques to be used. This review has discussed five areas of consideration that are often addressed in the available water sampling guidance: (1) CH/DH sample handling, (2) sampling bottle and equipment material and cleaning, (3) sampling device type and deployment, (4) filtration and preservation techniques appropriate for the analytical technique, and (5) QC samples to appropriately constrain sample process blanks. Below, we summarize the recommended sampling practices for routine total and dissolved metals (excluding Hg and sampling for chemical speciation), for which sub-ppm-level detection for all metals is required (see also Tier II sampling techniques in Figure 1):

1. Sample collection and handling should be performed using CH/DH techniques, with subsampling, processing, preservation, and other handling of open sample bottles ideally taking place in a Class 100 clean environment.
2. Sample bottles, if not provided by the laboratory performing the analyses, should be new, composed of uncolored plastic, stored in plastic bags, and acid-washed prior to sample collection.
3. Sampling equipment should be constructed of plastic or non-metal-containing materials and decontaminated, usually by cleaning with a dilute acid solution. Equipment should be stored in a clean area or in plastic bags prior to use for sampling.
4. Appropriate sample preservation often includes filtration and/or acidification. Filtration should occur soon after sampling, if required and possible. Routine trace metal samples should be acidified to pH <2, with the type and purity of acid dictated by the analytical method and the study's DQOs. The maximum holding time for samples to undergo analysis for most routine metals is 6 months.
5. QC samples must be collected to adequately quantify contamination associated with the sampling process and assess sampling and analytical bias (Wait et al., 2020). These generally include bottle blanks to assess whether the sample bottles need to be washed, field/trip blanks and equipment rinsate blanks to assess the effectiveness of the cleaning procedures, and filter and/or acid blanks to assess contamination associated with post-sampling preservation.

These protocols are sufficient for assessing most trace metal concentrations relative to the US EPA NRALWQ Criteria for the purposes of NRD Assessment.

A study's DQOs may require more or less stringent measures than those outlined above and/or deviations specific to samples intended for analyses other than total or dissolved trace metal concentrations, but these minimum precautions are considered sufficient for routine total and dissolved trace metal analyses at the ppb level (*i.e.*, Tier II Sampling Techniques). In some cases, the study's DQOs may not be well defined or may change during the course of the remedial activities (*e.g.*, between an initial remedial investigation and an ecological risk assessment performed at a later date). In these situations, it is advisable to choose sampling methodologies that will, at minimum, produce representative samples that can be measured for routine metal concentrations and be compared to the lowest metal assessment criterion of concern (*i.e.*, Tier II Sampling Techniques).

QC efforts should be continuous, and any change in sampling protocol should be documented. A change as simple as a batch of bottles from a different manufacturer can significantly alter blank levels. In some cases, however, contamination may be discovered during the course of sampling or after sampling is completed. While such a situation is not ideal, if sufficient QC samples have been collected, it may be possible to track the contamination to its source and determine whether it is consistent and well constrained. If consistent contamination can be identified and tracked *before* samples are taken, that contamination may be eliminated by increasing the level of precaution taken during the entire sampling process (*e.g.*, moving from the use of reagent-grade acid for bottle cleaning to trace metal-grade acid). If possible, working in a Class 100 clean laboratory space or glove box is generally encouraged, because contamination from the working environment is often random, inconsistent, and difficult to constrain. Ultimately, the appropriate sampling methodology is performance-based: even if the most protective sampling approach is employed, samples can still be contaminated if the sampling personnel are not sufficiently trained or if field conditions are not optimal. Factoring in time and effort for testing bottle blanks, analytical blanks, and other QC samples to assure that the sampling methodology is appropriately and efficiently suited to the DQOs and concentration range expected to be found in the field offers the best chance for a successful sampling campaign.

Funding

This research was supported by BP Exploration and Production Company.

References

American Society for Testing and Materials (ASTM). 2013. *Standard Practice for Sampling of Liquids in Waste Management Activities Using a Peristaltic Pump*. ASTM D 7353-07, 4p.

American Society for Testing and Materials (ASTM). 2016a. *Standard Practice for Sampling Liquids Using Grab and Discrete Depth Samplers*. ASTM D 6759-16, 9p.

American Society for Testing and Materials (ASTM). 2016b. *Standard Guide for Selection of Sampling Equipment for Waste and Contaminated Media Data Collection Activities*. ASTM D 6232-16, 39p.

Applied Marine Sciences, Inc. 2013. *Standard Operating Procedures: The Central Coast Areas of Special Biological Significance Regional Monitoring Program and Reference Site Monitoring (Version 1.1)*. Report to Monterey Regional Water Pollution Control Agency, 27p, May 15.

Arizona Department of Environmental Quality (ADEQ). 2012. *Standard Operating Procedures for Surface Water Quality Sampling*, 313p., September.

Arizona Dept. of Environmental Quality (ADEQ). 2015. *Standard Operating Procedures for Surface Water Quality Sampling*. Publication No. TB TB06-02, 345p., March.

Australia Northern Territory Government. 2009. *Methodology for the Sampling of Surface Water*. Dept. of Mines and Energy, Advisory # AA7-025 13p., March.

Australian and New Zealand Environment and Conservation Council (ANZECC); Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ) 2000. *Australian Guidelines for Water Quality Monitoring and Reporting*. National Water Quality Management Strategy No. 7. 251p., October.

Baltic Marine Environment Protection Commission (HELCOM). 2015. *Manual for Marine Monitoring in the COMBINE Programme of HELCOM*. 416p., February.

British Columbia Ministry of Water, Land and Air Protection. 2013. *British Columbia Field Sampling Manual For Continuous Monitoring and the Collection of Air, Air-Emission, Water, Wastewater, Soil, Sediment, and Biological Samples*. 401p., January.

California State Water Resources Board. 2007. *Marine Pollution Studies Laboratory – Department of Fish and Game (MPSL-DFG) Standard Operating Procedures (SOPs) for Conducting Field Measurements and Field Collections of Water and Bed Sediment Samples in the Surface Water Ambient Monitoring Program (SWAMP)*. Surface Water Ambient Monitoring Program (SWAMP), SOP Procedure No. 1.0, 64p., October 15.

Canadian Council of Ministers of the Environment. 2011. *Protocols Manual for Water Quality Sampling in Canada*. PN 1461, 186p.

Cutter, G. A., and Bruland, K. W. 2012. Rapid and noncontaminating sampling system for trace elements in global ocean surveys. *Limnology and Oceanography: Methods* 10(6):425–436. 216-6274

EPA Victoria. 2009. *Sampling and Analysis of Waters, Wastewaters, Soils and Wastes*. IWRG701. 36p. Accessed at http://www.epa.vic.gov.au/our-work/publications/publication/2009/june/iwrg701. 217-4129

Federal Advisory Committee on Detection and Quantitation Approaches and Uses in Clean Water Act Programs.

2007. *Report of the Federal Advisory Committee on Detection and Quantitation Approaches and Uses in Clean Water Act Programs (Final)*. Report to US EPA, Office of Water 176p., December 28.

Florida Dept. of Environmental Protection (FLDEP). 2014a. *FS 8200. Clean Sampling for Ultratrace Metals in Surface Waters (Revised)*. Water Quality Standards Program, EP-SOP-001/01. 14p., March 1. Accessed at http://www.dep.state.fl.us/water/sas/sop/sops.htm

Florida Dept. of Environmental Protection (FLDEP). 2014b. *FS 2000. General Aqueous Sampling*. DEP-SOP-001/01. 16p., March 1.

Gammon Technical Products, Inc. 2017. *Bacon Bomb Samplers*. 2p. Accessed at http://www.gammontech.com/mainframe/pdf/b115.pdf.

General Oceanics, Inc. 2016a. *Model 1010 Niskin bottle*.

General Oceanics, Inc. 2016b. *Model 1080 Go-Flo bottle*.

Georgia Dept. of Natural Resources (GADNR). 2008. *Waste Water Sampling*. Environmental Protection Division, SOP No. EPD-WPMP-5, 18p., January.

Gottler, R. A., and Piwoni, M. D. 2005. Part 3000: Metals. In *Standard Methods for the Examination of Water and Wastewater* (21st Ed.). American Public Health Association; American Water Works Association; Water Pollution Control Federation.

Government of Western Australia. 2009. *Surface Water Sampling Methods and Analysis – Technical Appendices: Standard Operating Procedures for Water Sampling – Methods and Analysis*. Department of Water, 66p., January.

International Organization for Standardization (ISO). 1992. *Water quality - Sampling - Part 9: Guidance on Sampling from Marine Waters*. ISO 5667-9:1992 (E), 16p.

Johnson, K. S., M., Gordon, R., and Coale, K. H. 1997. What controls dissolved iron concentrations in the world ocean? *Marine Chemistry* 57(3–4):137–161. 217-2061

Karin, R. W., Buono, J. A., and Fasching, J. L. 1975. Removal of trace elemental impurities from polyethylene by nitric acid. *Analytical Chemistry* 47(13):2296–2299.

Kneeland, J. M., Tcaciuc, A. P., Tuit, C. B., and Wait, A. D. 2020. A review of marine water sampling methods for hydrocarbons. Submitted.

Litman, R., Finston, H. L., and; Williams, E. T. 1975. Evaluation of sample pretreatments for mercury determinations. *Analytical Chemistry* 47(14):2364–2369.

Louisiana Dept. of Environmental Quality (LDEQ). 2008. *Standard Operating Procedure for Water Sample Collection, Preservation, Documentation and Shipping; Sonde Deployment and Continuous Monitoring (Revision 05)*. SOP_1134_R05, 59p., March 24.

Moll, M. [EMS Sistemas de Monitorizacion Medio Ambiental, SLU]. 2017. *Oil Spill Monitoring*. Accessed at http://www.seabird.com/oil-spill-monitoring.

Monterey Bay Aquarium Research Institute (MBARI). 2017. *Periodic Table of Elements in the Ocean*. Accessed at http://www.mbari.org/science/upper-ocean-systems/chemical-sensor-group/periodic-table-of-elements-in-the-ocean.

National Oceanic and Atmospheric Administration (NOAA) 1996. *Injury Assessment: Guidance Document for Natural Resource Damage Assessment Under the Oil Pollution Act of 1990*. 222p., August.

New Jersey Dept. of Environmental Protection (NJDEP). 2011. *Field Sampling Procedures Manual*. 574p., August.

Oklahoma Water Resources Board (OWRB). 2003. *Guidance Document for the Development of Site-Specific Water Quality Criteria for Metals*. OWRB Technical Report TRWQ2002-1, 91p.

Oklahoma Water Resources Board (OWRB). 2013. *Standard Operating Procedure for the Collection of Water Quality Samples in Lakes*. 16p., September.

Oregon Dept. of Environmental Quality (ORDEQ). 2013. *Field Sampling Reference Guide*. Laboratory and Environmental Assessment Division, DEQ86-LAB-0002-QAG Version 8.0, 95p., July 5.

Parker, S. P., ed. 1984. Surface water. In *McGraw-Hill Dictionary of Scientific and Technical Terms* (3rd Ed.). New York: McGraw-Hill Book Co., 1589.

Patterson, C., Settle, D., and Glover, B. 1976. Analysis of lead in polluted coastal seawater. " *Marine Chemistry* 4(4):305–319.

Patterson, C. C., and Settle, D. M. 1976. The reduction of orders of magnitude errors in lead analysis of biological materials and natural waters by evaluating and controlling the extent and sources of industrial lead contamination introduced during sample collection, handling and analysis. In *Accuracy in Trace Analysis; Sampling, Sample Handling Analysis, Volume 1*, ed. LeFleur, P. D. U.S. Government Printing Offices, Washington, D.C. US Dept. of Commerce, National Bureau of Standards Special Publication, 422, 321–351.

Research Planning, Inc.; National Oceanic and Atmospheric Administration (NOAA). 2014. *Guidelines for Collecting High Priority Ephemeral Data for Oil Spills in the Arctic in Support of Natural Resource Damage Assessments (Final)*. 284p., September.

Sampling Systems Ltd 2017. *TeleScoop*. Accessed at http://www.sampling.com/TeleScoop.html.

Sea-Bird Scientific. 2017. *SBE 32 Carousel Water Sampler*. Accessed at http://www.seabird.com/sbe32-carousel-water-sampler#32fullsize.

Tcaciuc, A. P., Kneeland, J. M., Tuit, C. B., and Wait, A. D. 2020. A review of marine oil sampling methods. Submitted.

Texas Commission on Environmental Quality (TCEQ). 2012. *Surface Water Quality Monitoring Procedures, Volume 1: Physical and Chemical Monitoring Methods*. RG-415, 202p., August.

Thiers, R. E. 1957. Contamination in trace analysis and its control. In. *Methods of Biochemical Analysis*, Vol. V, ed. Glick, D. New York: Interscience Publishers, Inc, 273–335.

Tuit, C. B., Ritter, H. C., and Wait, A. D. 2020. A review of marine sediment sampling methods. Submitted.

US Dept. of Agriculture (USDA). 2001. *Field Guide for Surface Water Sample and Data Collection*. Forest Service Air Program, 73p., June.

US EPA Region 1. 2003. *Standard Operating Procedure for the Collection of Low Level Metals Ambient Water Samples (Revision No. 1)*. Office of Environmental Measurement and Evaluation, 12p., September 4.

US EPA Region 10; Puget Sound Water Quality Action Team. 1997. *Recommended Guidelines for Sampling*

Marine Sediment, Water Column, and Tissue in Puget Sound. 59p., April.
US EPA Region 4. 2013b. *Field Sampling Quality Control (Operating Procedure)*. SESDPROC-011-R4, 25p.
US EPA Region 6. 2000. *Collecting Water-Quality Samples for Dissolved Metals-In-Water (Revised)*. 20p.
US EPA, Region 4. 2013a. *Surface Water Sampling*. Science and Ecosystem Support Division, SESDPROC-201-R3. 22p, February 28.
US EPA. 1982. *Handbook for Sampling and Sample Preservation of Water and Wastewater*. Environmental Monitoring and Support Laboratory, EPA-600/4-82-029, 418p., September.
US EPA. 1986. *Test Methods for Evaluating Solid Waste: Laboratory Manual, Physical/Chemical Methods, Volumes 1A, 1B and 1C [SW-846]*. Office of Solid Waste, EPA SW-846, November.
US EPA. 1994. *Surface Water Sampling*. SOP # 2013. Accessed at www.dem.ri.gov/pubs/sops/wmsr2013.pdf, 7p., November 17.
US EPA. 1996a. *Method 1669: Sampling Ambient Water for Trace Metals at EPA Water Quality Criteria Levels*. Office of Water, EPA 821-R-96-011, 33p., July.
US EPA. 1996b. *Method 1638:* Determination of Trace Elements in Ambient Waters by Inductively Coupled Plasma – Mass Spectrometry. 50p., January.
US EPA. 1996c. *Method 1640: Determination of Trace Elements in Ambient Waters by On-Line Chelation Preconcentration and Inductively Coupled Plasma-Mass Spectrometry*. Office of Water, 47p., January.
US EPA. 2004. *National Recommended Water Quality Criteria*. Office of Water; US EPA, Office of Science and Technology, 25p. Accessed at http://www.epa.gov/water-science/criteria/wqctable/nrwqc-2004.pdf.
US EPA. 2009. *Guidance for Implementing the January 2001 Methylmercury Water Quality Criterion (Final)*. Office of Science and Technology, EPA 823-R-09-002, 211p., January.
US EPA. 2016a. *Draft Aquatic Life Ambient Estuarine/Marine Water Quality Criteria for Copper - 2016*. Office of Water, EPA-822-P-16-001, 257p., July.
US EPA. 2016b. *Definition and Procedure for the Determination of the Method Detection Limit, Revision 2*. Office of Water, EPA-821-R-16-006, 8p., December.
US EPA. 2016c. *Aquatic Life Ambient Water Quality Criterion for Selenium – Freshwater (2016)*. Office of Water, EPA 822-R-16-006, 807p., June.
US EPA. 2017. *National Recommended Water Quality Criteria - Aquatic Life Criteria Table*. Accessed at https://www.epa.gov/wqc/national-recommended-water-quality-criteria-aquatic-life-criteria-table, May 9.
US EPA; US Army Corps of Engineers (US ACE). 1995. *QA/QC Guidance for Sampling and Analysis of Sediments, Water, and Tissues for Dredged Material Evaluations: Chemical Evaluations*. EPA 823-B-95-001, 332p., April.
US Geological Survey (USGS). 2011. *U.S. Geological Survey Protocol for Sample Collection in Response to the Deepwater Horizon Oil Spill, Gulf of Mexico, 2010: Sampling Methods for Water, Sediment, Benthic Invertebrates, and Microorganisms in Coastal Environments*. Open-File Report 2011–1098. Accessed at http://pubs.usgs.gov/of/2011/1098.
US Geological Survey (USGS). 2012. *Groundwater Quality in West Virginia, 1993–2008*. USGS Scientific Investigations Report 2012–5186, 47p.
US Geological Survey (USGS) 2015. *National Field Manual for the Collection of Water-Quality Data: Techniques of Water-Resources Investigations (Book 9)*. 1539p., October.
Utah Dept. of Environmental Quality (UTDEQ). 2011. *Standard Operating Procedure for Collection of Water Chemistry Samples, GSL Impounded Wetlands 2012 Monitoring Activities (Revision 1)*. Division of Water Quality, 23p., September 10.
Virginia Dept. of Environmental Quality (VADEQ). 1996. *Memorandum to Regional Directors re: Guidance Memorandum No. 96-009; Obtaining Dissolved Metals Data*. Water Division, 37p., November 19.
Wait, A. D., Tuit, C. B., and Maney, J. P. 2020. Forensic sampling practices for oil spills in the marine environment.
Washington State Dept. of Ecology (WADOE). 2010. *Standard Operating Procedure for the Collection and Field Processing of Metal Samples*. 12p., October.
Wildco. 2016a. *Kemmerer Bottle*.
Wildco. 2016b. *Alpha Bottle Only - 3.2L Vertical, PVC*. Accessed at http://shop.sciencefirst.com/wildco/alpha-water-samplers-vertical/6723-alpha-water-sampler-vertical-pvc-water-sampler-only-opaque-pvc-32l.html.