# The Montana Department of Environmental Quality Metals Assessment Method

**Final**

**July 2012**

**Prepared by:**
Jonathan Drygas
Water Quality Planning Bureau, Monitoring and Assessment Section
Montana Department of Environmental Quality
1520 E. Sixth Avenue
P.O. Box 200901
Helena, MT 59620-0901

WQPBMASTR-03

**Suggested citation:** Drygas, Jonathan. 2012. The Montana Department of Environmental Quality Metals Assessment Method. Helena, MT: Montana Dept. of Environmental Quality.

# TABLE OF CONTENTS



# LIST OF TABLES



# LIST OF FIGURES

# ACRONYMS

| **Acronym** | **Definition** |
|---|---|
| ADB | Assessment database |
| CEPA | California Environmental Protection Agency |
| DEQ | Department of Environmental Quality (Montana) |
| EPA | Environmental Protection Agency (US) |
| MDL | Method Detection Limit |
| RL | Reporting Limit |
| SOP | Standard Operating Procedures |
| TMDL | Total Maximum Daily Load |
| WQS | Water Quality Standards |

# 1.0 Introduction

Metals pollutants can adversely affect the beneficial uses for aquatic life/fish and human health. The purpose of this document is to provide an assessment method and monitoring framework to determine attainment of numeric metals water quality standards as defined in Circular DEQ-7 (Montana Department of Environmental Quality, 2010).

# 2.0 Sampling Frame

The metals assessment method applies to all Montana surface waters. For lakes and reservoirs, the sampling frame is applicable to lakes greater than or equal to 1 acre. Lake sampling is conducted at a single location (mid-depth lake), although additional sites may be included to address risk-based sampling (i.e., bays). For streams and rivers, the sampling frame is a stream segment. Sampling frames and stratification procedures for rivers and streams are similar to those described for wadeable streams in **Section 2.1**.

## 2.1 Sample Frame, Population, and Sampling Units for Wadeable Streams

This section follows similar guidelines as Appendix A of the “Assessment Methodology for Determining Wadeable Stream Impairment Due to Excess Nitrogen and Phosphorus Levels”(Suplee and Sada de Suplee, 2011). It is important to note that an assessment reach for nutrients may not always be the same for metals but the sampling frame concept for wadeable streams is the same.

All studies involving statistical evaluations of data require that a sample frame, population, and sampling unit be defined. For the purposes of determining compliance with numeric metals water quality standards, we define the following:

- **Sample Frame**: *A segment listed in the Assessment Data Base (ADB)* ***OR*** *a sub-segment of an ADB stream segment.*
- **Population**: *All the water flowing through the segment.*
- **Sampling Unit:** A sample collected from the segment that is largely independent of other samples collected within the segment.

Each segment will be made up of a series of reaches and sampling sites (**Figure 2-1**). Reaches are subdivisions of segments that represent different extents based on geomorphology, land use, or other peripheral influences including adits and tributaries. Segments that are homogeneous throughout their entire length may be considered a single reach while segments with a combination of changes in geomorphology, land use and influences may result in multiple reaches. Sampling sites are selected by the assessor to represent a portion of the reach. The assessors will thoroughly review the stream segment to develop an economical and comprehensive sampling design to adequately characterize the stream segment. This includes the determination of representative sampling sites and their accessibility. The data required for these considerations includes historical and current land use information, maps and landowner information (Montana Department of Environmental Quality, 2005).

*Assumptions:* For purposes of determining compliance with numeric metals standards, it is usually assumed that (1) pollution sources are evenly dispersed along the reach, (2) sampling sites are randomly located along the reach, and (3) each sample is independent of the others. Spatial and temporal independence guidelines for sites are addressed in **Section 2.2** below.

In some cases, ADB segments may have area of higher pollution concentrated only in a particular part of the stream (hotspots). In such cases, it may not make sense to view the original ADB segment as the best possible sampling frame. That is, it would be better to further stratify the sample frame and, thus, the population of interest. This will prevent statistical distortion of results caused by mixing data together, for common analysis, data from the relatively un-impacted sub-segment with data from the impacted sub-segment. For example, in **Figure 2-1** it might be prudent to consider the sub-segment upstream of the Star Mine as a sampling frame apart from the sub-segment below the mine. Stratification is common in studies employing purely random sampling, where it is referred to as stratified random sampling (Cochran, 1977). Stratification allows maximal precision of estimates with minimal sampling effort (Norris et al., 1992). The assessor carrying out the analysis on an ADB segment will have to judge if further stratification is warranted. If it is warranted, then sampling requirements, described above and further detailed below, would apply to *each* of the new sub-segments individually.

*Precautionary Considerations*: Pollution sources are rarely evenly-dispersed along stream segments, violating assumption 1 above. Purely random sampling is usually not practical due to stream access issues, etc. Targeting only the known or potential pollution “hotspots” — even within an assessment reach that has been broken out from a larger ADB segment — could over represent the hotspots and distort the statistical tests. Sampling and analysis plans (SAPs) should proceed with goal-oriented sampling (Buck et al., 2000) that works towards striking an equitable balance between the number of hotspot sites and the number of un- or minimally impacted sites *within the defined assessment reach*. That is, the aggregate of collected samples should be representative(Environmental Research Laboratory-Duluth, 2002) of the assessment reach as a whole. Advanced knowledge and expertise of the field will be needed to accomplish this (Norris et al., 1992), and modifications to the assessment reach boundaries can be made on-the-fly during field work, if deemed necessary. It is possible to sub-segment a stream reach to the point where, for a particular assessment reach, there really is little left but hotspots; if this is the case, and the assessor believes it is appropriate, then the hotspots *are* representative of the assessment reach. As a general rule, it is better to lump than split to avoid unnecessary sampling and administrative work. The requirement to create reasonably uniform assessment reaches is inherently in conflict with the need to “lump” for the purpose of keeping assessments as simple as possible. Judgment is needed to balance these two opposed factors and come up with an optimal sampling strategy.

Although this quasi-systematic approach is not a substitute for truly random sampling it will, if carried out properly, achieve good sample interspersion and representativeness. For further discussion of randomization *vs.* interspersion approaches, see page 196 of (Hurlbert, 1984).

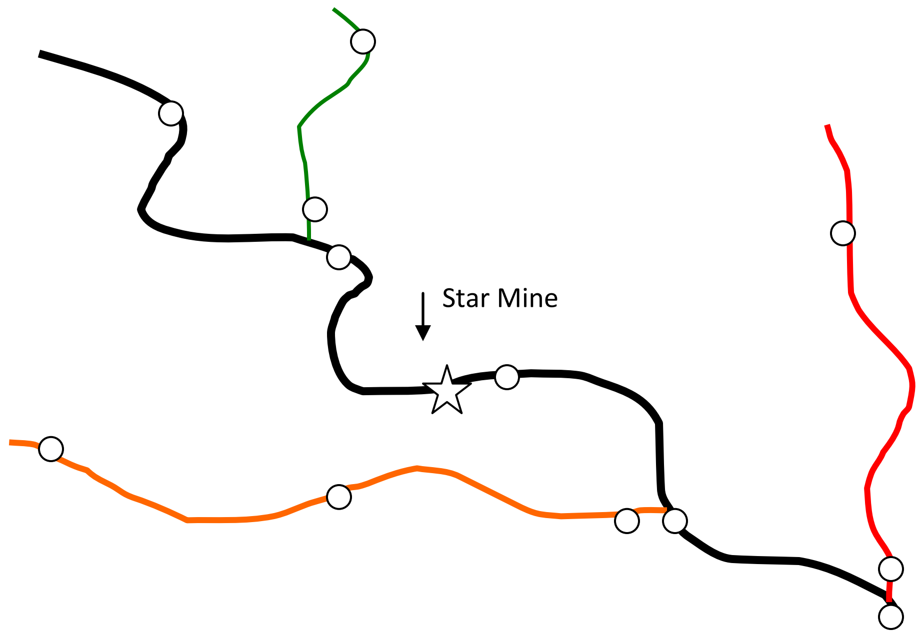


**Figure 2-1. Four different stream reaches (shown by different colors), each representing 1 sampling frame (ADB stream segment).**
Example sampling sites (hollow dots) are shown along each segment.

## 2.2 Sampling Independence

Sampling independence includes spatial and temporal components. Samples can be collected any time of the year. Ideally, 33% of the data set should be from samples collected during high flow conditions (i.e., spring runoff periods when streams are reaching maximum flow and expected pollutant loads are likely mobilized), while the remaining should be collected during baseflow conditions. Samples that are all collected during a short-term event (e.g., storm, flood) would not be considered independent and representative of true conditions. Therefore, a sampling design ratio of one high flow sample per every two baseflow sampling events would provide an unbiased data set.

During high flow, the temporal component will be evaluated on a case by case basis due to the changing discharge conditions. During baseflow, there should be a 7-day period between sampling events.

The spatial component is set at 1 mile between sampling sites for streams. If sources for metals (e.g. tailings piles, abandoned or active mines, etc.) are identified, bracketing the sample sites in relation to the source should occur, and in some cases, the 1 mile spatial component might not apply.

# 3.0 Assessment Methodology

When assessing the metals data, both the aquatic life/fish and drinking water beneficial uses should be evaluated. Numeric standards to protect aquatic life and human health are different and therefore the methods on how they are applied differ (Montana Department of Environmental Quality, 2010). In general, some standard exceedances are allowed to assess the aquatic life/fish beneficial use, with the exception of silver, which is interpreted as a “not to exceed” value. No exceedances are allowed for the drinking water beneficial use (human health standard). Thus, a waterbody can be listed for not supporting the drinking water beneficial use but not for the aquatic life/fish beneficial use. In addition, it’s important to note that some of the aquatic life standards are dependent on hardness and adjust with changes to the hardness. Hardness values from the same sampling event are required for the assessment of hardness dependent metals.

The minimum sample size is 8 independent samples within the same assessment frame. Fewer samples may be considered (see specifics for each beneficial use attainment decision). In general, data from the last **ten** years should be only considered when making attainment decisions for both beneficial uses (aquatic life/fish and drinking water). However, if conditions are known not to have changed, DEQ would consider including data older than 10 years in the analysis if no more recent data is available. Data that is older than ten years can be used as a historical reference and for Total Maximum Daily Load (TMDL) development.

At a minimum, for non targeted sampling designs, a metals sampling suite should include the total recoverable metals, dissolved aluminum and sediment metals listed in **Table 3-1.** Sediment metals data are not used for making assessment decisions but are valuable in determining potential sources. The rationale for using this subset of metals is based on the most commonly 303(d) listed metals in Montana and their metals associations. DEQ analyzed metals associations in order to minimize the cost of monitoring. The analysis showed that some metals had strong associations with a metal included in the basic monitoring suite, and did not need to be monitored. If these excluded metals are known to be an issue (e.g., a known nickel source exists or is already listed as a cause of impairment) they will be added to the monitoring suite on a case-by-case basis. The specific project needs will determine if additional parameters and sample fractions are necessary.

**Table 3-1. Recommended minimal metals sampling suite.**

| **Water Sample - Total Recoverable Metals** | | |
|---|---|---|
| **Metal** | **Method** | **Req. Report Limit (ug/L)** |
| Arsenic | EPA 200.8 | 1 |
| Cadmium | EPA 200.8 | 0.03 |
| Chromium | EPA 200.8 | 10 |
| Copper | EPA 200.8 | 2 |
| Iron | EPA 200.7 | 20 |
| Lead | EPA 200.8 | 0.3 |
| Selenium | EPA 200.8 | 1 |
| Silver | EPA 200.8 | 0.2 |
| Zinc | EPA 200.7 | 8 |
| Calcium (for calculating hardness) | EPA 200.7 | 1000 |
| Magnesium (for calculating hardness) | EPA 200.7 | 1000 |
| Total Hardness | A2340B (calculated) | 1000 |
| **Water Sample - Dissolved Metals** | | |
| Aluminum | EPA 200.7 | 9 |
| **Sediment Sample – Metals** | | |
| **Metal** | **Method** | **Req. Report Limit (mg/kg - dry weight)** |

| | | |
|---|---|---|
| Arsenic | EPA 200.8 | 1 |
| Cadmium | EPA 200.8 | 0.2 |
| Chromium | EPA 200.8 | 9 |
| Copper | EPA 200.8 | 15 |
| Iron | EPA 200.7 | 10 |
| Lead | EPA 200.8 | 5 |
| Zinc | EPA 200.7 | 20 |
| Mercury | EPA 7471B | 0.05 |

## 3.1 Aquatic Life/Fish Beneficial Use

Numeric standards for aquatic life support are outlined in the Circular DEQ-7 (Montana Department of Environmental Quality, 2010). Aquatic life standards are for both acute and chronic exposure. Acute criteria are based on a one hour exposure event and can only be exceeded once, on average, in a three year period. Chronic criteria are based on a 96 hour exposure and can only be exceeded, on average, once on a three year period.

There are two parts of the standards that need discussion: a) the frequency component of the standard of one exceedance in three years (for both acute and chronic criteria), and b) the duration component of the standard of 96 hour exposure (for chronic criteria only).

a) One exceedance in three years: DEQ needed to develop a method to address this frequency component of the standard. The difficulty with this approach is that attainment decisions do not take into consideration the number of results. The exceedance rate for attainment decisions will vary with the size of the data set. DEQ evaluated exceedance rates and the number of samples that would be required to achieve an accepatble level of certainty to determine what is realistically affordable. Thus,
   - Annual sampling- 3 samples over three years, one exceedance allowed: rate = 0.333
   - Semi-annual sampling- 6 samples over three years, one exceedance allowed: rate = 0.166
   - Quarterly sampling- 12 samples over three years, one exceedance allowed: rate = 0.0833
   - Monthly sampling- 36 samples over three years, one exceedance allowed: rate = 0.0277

   As the data set increases, the allowable exceedance rate is reduced and the potential for attainment decision errors changes. In looking at these four scenarios, the two extremes (annual sampling and monthly sampling) are not realistic due to a lack of confidence (just having 3 samples) or due to the cost (36 samples). Often, DEQ has limited sample sizes since monitoring in a state the size of Montana is so resource intensive, so the sampling scenarios of semi-annual and quarterly sampling are most representative of the size of typical data sets.

   When determining the allowable exceedance rate, DEQ begun its analysis by determining the level of sampling it could afford for metals. Funding limitations placed the number of samples that could be afforded to a range of 8-14 samples per waterbody segment. DEQ then used a cumulative binomial distribution to back calculate the confidence based on the affordable sample size. An exceedance rate of 10% was selected as multiple states have set their exceedance rate at 10% and this rate has also been recommended in other state's guidance documents (California Environmental Protection Agency State Water Resources Control Board, 2004). Using each value between 8 and 14, inclusive, DEQ applied the exceedance rate historically used of 10% and calculated the certainty that would result. The resulting 65% certainty of making the right call was acceptable given the economic constraints and DEQ determined that the 10% exceedance rate balances the confidence level with the cost of sampling.

DEQ was also informed by the analysis performed by California Environmental Protection Agency (CEPA) in their development of the Functional Equivalent Document, Water Quality Control Policy for Developing California's Clean Water Act 303(d) List (California Environmental Protection Agency State Water Resources Control Board, 2004). In their analysis, CEPA reviewed many alternatives for statistical analysis and determined the cost for the statistical alternatives they had selected. In that analysis, CEPA determined a minimum sample size of 22 samples for metals analysis at a 10% exceedance rate using unequal alphas in binomial distribution tests that were set to two different null hypotheses depending upon the water's current status ($H_o$ water impaired for listed waters, $H_o$ water unimpaired for unlisted or unknown waters). Rather than going through a similar exercise only to find that the level of statistical certainty desired was out of reach financially, DEQ a cumulative binomial distribution to calculate the error probability based on an affordable level of data.

b) 96 hour exposure: Circular DEQ-7 (Montana Department of Environmental Quality, 2010) states the chronic standard is based on a 96 hour exposure. Typically, the assessor will have just one sample from a single site in any given 96 hour period, so a single sample may be used to represent the 96 hour average. In the event that there are multiple samples from the same site during a 96 hour period, then the results will be averaged together.

For aquatic life/fish assessment, a minimum of 8 independent samples should be collected from the same assessment reach. The maximum number of samples to be collected for this assessment method is 20 samples. The maximum sample size is used as an end point so the assessor is not in an endless loop of sampling when dealing with borderline exceeding datasets. By setting a maximum sample size of 20, the alpha and beta error rate is about 35% (Drygas, Jonathan and Mark Bostrom, personal communication 2011). If the alpha and beta error rate was reduced to 25%, then 50 samples would be the maximum. Selecting 35% is realistic for the size of Montana and the number of waters. If more than 20 samples have already been collected, the most current data should be considered when applying the assessment method. A method for how to select independent samples and deal with larger data sets samples is in development and will be addressed at a future date.

In some cases the data sets may be small and additional sampling is not feasible. Rigorous data collection is unnecessary when there are a minimum of 2 exceedances of aquatic life standards in data sets of 3-7 samples. In these situations, if sources are identified, then the attainment decision will be to list or to remain listed.

If more than 10% of the samples exceed the standard, then the attainment decision is to list or to remain listed. If the exceedance rate is equal to or less than 10%, then the attainment decision is not to list or delist. There are three exceptions to the 10% exceedance rate attainment decisions: a) silver must not exceed the acute standard; b) if **twice** the **acute** standard is exceeded in a sample, then the attainment decision is to list or to remain listed regardless of the percent exceedance by the data set or the data set size; c) if the 10% exceedance rate threshold is surpassed but no human caused metals sources are found in the drainage, then the assessor should consult management for a case by case review.

## 3.2 Drinking Water Beneficial Use (Human Health Standard)

Numeric standards for human health are outlined in the Circular DEQ-7 (Montana Department of Environmental Quality, 2010). No standard exceedances are allowed when assessing for human health. A minimum sample size for an assessment reach is eight samples. However, if just one sample exceeds the human health standard, then the attainment decision is to list or to remain listed. If there are no exceedances of human health standards then the attainment decision should be not to list or to delist.

## 3.3 Non-detects and J Flagged Samples

The use of non-detect data points is dependent on the relation of the Water Quality Standard (WQS) to the detection limit. If the WQS is a higher value than the detection limits, then the non-detect data point can be used in the data set. If a data point is reported as non-detect and the associated detection limit is higher than the WQS, then it becomes uncertain if the standard is exceeded and the data point should not to be used as part of the data set.

Data flagged with a "J" are to be handled in a similar manner. Data are "J" flagged when the analytical result falls between the Reporting Limit (RL) and the Method Detection Limit (MDL) but the value given is considered an estimate with a low error bar. The use of "J" flagged data is dependent on where the WQS is in relation to the reporting limit and the method detection limit (**Figure 3-1**).

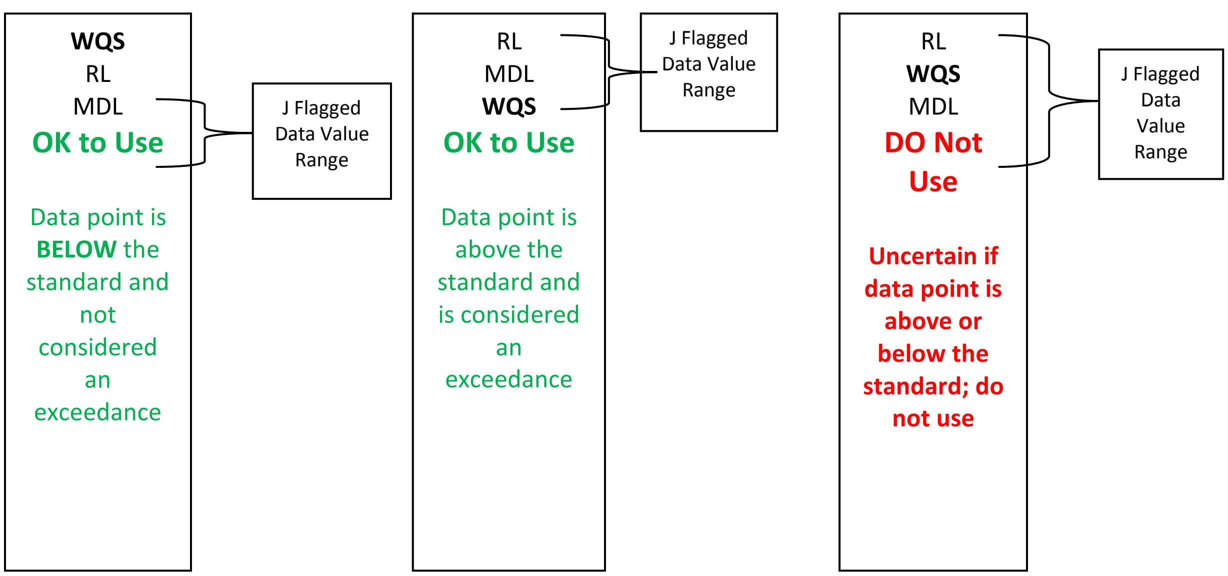


**Figure 3-1. Shows three scenarios involving J flagged data and the Water Quality Standard.**

The figures show that two of three possibilities result in usable data. (1) When the reporting limit is below the WQS, the J-flagged data point can be used and is not counted as an exceedance. (2) If the reporting limit and method detection limit are above the WQS, the J-flagged data point can be used and it should be counted as an exceedance. (3) Only when the standard is between the method detection limit and the reporting limit, the J-flagged data point becomes unusable. This is due to the uncertainty if the data value is above or below the standard and the value should not be used in making an impairment decision.

# 4.0 References


Buck, Sharon, Walter K. Dodds, Jen Fisher, David A. Flemer, Debra Hart, Amanda K. Parker, Jan Stevenson, Vicki Watson, and Eugene B. Welch. 2000. Nutrient Criteria Technical Guidance Manual, Rivers and Streams. Washington, DC: United States Environmental Protection Agency. Report EPA-822-B00-002. http://www.epa.gov/waterscience/criteria/nutrient/guidance/rivers/index.html.

California Environmental Protection Agency State Water Resources Control Board. 2004. Functional Equivalent Document, Water Quality Control Policy for Developing California's Clean Water Act Section 303(d) List -Final. California: **California Environmental Protection Agency State Water Resources Control Board**.

Cochran, William G. 1977. Sampling Techniques, 3 ed., New York: John Wiley and Sons.

Drygas, Jonathan and Mark Bostrom. 2011. Personal Communication.

Environmental Research Laboratory-Duluth. 2002. Consolidated Assessment and Listing Methodology: Towards a Compendium of Best Practices. Washington, D.C.: U.S. Envrionmental Protection Agency.

Hurlbert, Stuart H. 1984. Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs.* 54(2): 187-211.

Montana Department of Environmental Quality. 2005. Field Procedures Manual For Water Quality Assessment Monitoring. Helena, MT: Montana Department of Environmental Quality, Water Quality Planning Bureau. Report WQPBWQM-020.

-----. 2010. Circular DEQ-7: Montana Numeric Water Quality Standards. Helena, MT: Montana Department of Environmental Quality. http://deq.mt.gov/wqinfo/Standards/PDF/DEQ-7.pdf. Accessed 6/9/2011.

Norris, Richard H., E. P. McElravy, and V. H. Resh. 1992. "The River Handbook,", Calow, Peter and Petts, Jeffrey E., Ch. The Sampling Problem, (Oxford, England: Blackwell Scientific Publications)

Suplee, Michael W. and Rosie Sada de Suplee. 2011. Assessment Methodology for Determing Wadeable Stream Impairment Due to Excess Nitrogen and Phosphorus Levels. Helena, MT: Montana Department of Environmental Quality.