Derivation of the Nonanthropogenic Arsenic Standards for Segments of the Upper and Middle Yellowstone River

October 2019

Prepared by:
Water Quality Standards & Modeling Section
Montana Department of Environmental Quality
1520 E. Sixth Avenue
P.O. Box 200901
Helena, MT 59620-0901

Suggested citation: Montana Department of Environmental Quality. 2019. Derivation of the Nonanthropogenic Arsenic Standards for Segments of the Upper and Middle Yellowstone River. Helena, MT: Montana Dept. of Environmental Quality.

EXECUTIVE SUMMARY

Arsenic concentrations along much of the Yellowstone River are consistently above Montana's human health standard of 10 µg/L. Per Montana law, it is not necessary to treat wastes to a condition purer than the natural condition (75-5-306, MCA). Similarly, the Department of Environmental Quality (DEQ) may not apply a water quality standard to a water body that has a nonanthropogenic concentration greater than the standard (75-5-222, MCA). 75-5-222, MCA goes on to say that, in such cases, the nonanthropogenic concentration *is* the standard.

In the demonstration of nonanthropogenic arsenic (DON) report (DEQ, 2019), DEQ presented the methods and results for demonstrating the nonanthropogenic proportion of arsenic in the Yellowstone River study region. The study region includes the Yellowstone River from the Montana/ Wyoming Border to the mouth of the Bighorn River near Bighorn, Montana. The river was divided into five hydrologic segments in the DON report as well as in this document:

- Segment 1 - Montana/Wyoming border to the mouth of Mill Creek near Pray, MT
- Segment 2 - Mill Creek to the mouth of the Boulder River near Big Timber, MT
- Segment 3 - Boulder River to the mouth of the Stillwater River
- Segment 4 - Stillwater River to the mouth of the Clarks Fork of the Yellowstone River
- Segment 5 - Clarks Fork of the Yellowstone River to the mouth of the Bighorn River

In the DON report, DEQ determined that the nonanthropogenic concentrations and percentages of arsenic are highest in Segment 1 immediately downstream from Yellowstone National Park (YNP). Arsenic concentrations then notably decrease in the Yellowstone River in a downstream direction; still, the proportion of the arsenic load attributable to YNP remains above 90% throughout the study region. Annually, 97% of the total arsenic load in the Yellowstone River is nonanthropogenic at the Montana/Wyoming border, while at the mouth of the Bighorn River it is 96%.

DEQ determined that the Yellowstone River has a high flow season from May 1 to July 31 and a low flow season from August 1 to April 30; Yellowstone River arsenic concentrations were shown to vary by these high and low flow seasons. As a result, DEQ has identified unique arsenic standards applicable to each season.

Based on the results of the DON report, Yellowstone River nonanthropogenic arsenic concentrations are identified in this document by segment and season and shown in the table below. DEQ recommends that all the values in the table be adopted as nonanthropogenic standards except for the Segment 5 high flow value. In Segment 5 during high flow, the river's median nonanthropogenic concentration is lower than the existing human health standard of 10 µg/L, and DEQ does not recommend changing the currently-adopted human-health based standard in such situations.

Yellowstone River Segments and their Median Nonanthropogenic Arsenic Concentrations

Segment	Description	High Flow Season Arsenic Concentration[1] (µg/L)	Low Flow Season Arsenic Concentration[1] (µg/L)
1	MT/WY Border to Mill Creek	11	32
2	Mill Creek to Boulder River	11	24
3	Boulder River to Stillwater River	10	18
4	Stillwater River to Clarks Fork Yellowstone	10	14
5	Clarks Fork Yellowstone to Bighorn River	8	10

[1] High Flow season for the Yellowstone River is May – July, and the Low Flow season is August - April.

Implementation of the standards, for purposes of both ambient river assessment and discharge permitting, are discussed in the last section of this report.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

ARM	Administrative Rules of Montana
cfs	cubic feet per second
DEQ	Department of Environmental Quality
DON	Demonstration of Nonanthropogenic Condition
EPA	Environmental Protection Agency
HHS	Human Health Standard
kg/day	kilograms per day
LCL	Lower Confidence Level
MCA	Montana Code Annotated
MPDES	Montana Pollutant Discharge Elimination System
NAS	Nonanthropogenic Arsenic Standard
NSE	Nash-Sutcliffe Coefficient of Efficiency
QAPP	Quality Assurance Project Plan
SAP	Sampling and Analysis Plan
TMDL	Total Maximum Daily Load
µg/L	micrograms per liter
USGS	United States Geological Survey
WQPB	Water Quality Planning Bureau
WQSM	Water Quality Standards and Modeling Section
YNP	Yellowstone National Park

1.0 INTRODUCTION

This document presents the methods the Department of Environmental Quality (DEQ) used to derive arsenic standards for portions of the Yellowstone River based on the river's natural (nonanthropogenic) arsenic concentrations. Collectively, this work is referred to here as the nonanthropogenic arsenic standard (NAS). The work was completed by the Water Quality Standards & Modeling Section (WQSM) within the Water Quality Planning Bureau (WQPB). Assistance was also provided by DEQ's Water Protection Bureau. The geographic area encompassed in this document includes the Yellowstone River from the Montana/Wyoming Border to the mouth of the Bighorn River near Bighorn, Montana.

1.1 PURPOSE

The purpose of this NAS document is to develop appropriate arsenic water quality standards for sections of the Yellowstone River where arsenic concentrations are equal to or above the adopted human health standard (HHS) of 10 µg/L due to nonanthropogenic sources. Per Montana law, DEQ may not apply a water quality standard to a water body that has a nonanthropogenic concentration greater than the standard (75-5-222, MCA). The same law goes on to say that, in such cases, the nonanthropogenic concentration is the standard. Also, Montana law has stated since 1967 that dischargers are not required to discharge to purer than natural (75-5-306, MCA).

1.2 SUPPORTING DOCUMENTS

The nonanthropogenic portion of the total arsenic load for the Yellowstone River was identified in the document *Demonstration of Nonanthropogenic Arsenic Levels: Yellowstone River, Montana* (DON; DEQ, 2019). That document summarized the methods and results for determining seasonal nonanthropogenic arsenic loads and the percentage of the total arsenic load that was nonanthropogenic in the Yellowstone River. The term "DON" will be referenced throughout this document, and the term is used interchangeably with the citation DEQ, 2019.

The quality assurance descriptions for field data collection, data compilation, and modeling described in this document were provided in the *DEQ Quality Assurance Project Plan* (QAPP) and *Sampling and Analysis Plans* (SAP; DEQ, 2015a; 2015b; 2016a; 2017b).

1.3 BACKGROUND

The Yellowstone Caldera within Yellowstone National Park (YNP) is the largest source of arsenic to the Yellowstone River (YNP, 2015). Due to this geothermal activity, a large stretch of the Yellowstone River has arsenic concentrations elevated above Montana's HHS of 10 µg/L (Circular DEQ-7, June 2019).

DEQ demonstrated that the nonanthropogenic concentrations and percentages of arsenic are highest in Segment 1 immediately downstream of Yellowstone National Park, and decrease as the Yellowstone River travels downstream (DEQ, 2019). Annually, 97% of the total arsenic load in the Yellowstone River is nonanthropogenic at the Montana/Wyoming border, decreasing to 96% at the mouth of the Bighorn River. In this document, DEQ will develop the Yellowstone River nonanthropogenic standards for arsenic based on the results from DEQ (2019).

1.3.1 Montana Water Quality Standards

Montana water quality standards are maintained in ARM 17.30 Subchapter 6. The reach of the Yellowstone River addressed in this document spans three different use classes:

- B-1: from the Montana/Wyoming border to the Laurel water supply intake;
- B-2: from the Laurel water supply intake to the Billings water supply intake; and
- B-3: from the Billings water supply intake to the Bighorn River.

All three of these “B” classes have the same beneficial use (drinking water) which has the potential to be impacted by elevated arsenic concentrations. State law requires that waterbodies in these B classes be maintained suitable for drinking purposes after conventional treatment. As of this writing, the numeric water quality standard in Circular DEQ-7 for arsenic is 10 µg/L, and is applicable to all three B classes. In DEQ-7, arsenic is categorized as a human health carcinogen. Surface water arsenic concentrations and standards refer to total recoverable arsenic.

1.3.2 Yellowstone River Arsenic Segment Delineation

In this document, DEQ will identify nonanthropogenic arsenic standards which can supersede the arsenic HHS for those stretches of the Yellowstone River that have median (i.e., 50^{th} percentile) nonanthropogenic arsenic concentrations equal to or above the 10 µg/L HHS.

DEQ determined the scope of the project by identifying the extent of the Yellowstone River that had arsenic levels above 10 µg/L for at least part of the year (DEQ, 2019). The project reach was determined to be the Yellowstone River from the Montana/Wyoming border to the mouth of the Bighorn River. This reach was divided into five hydrologic segments as presented in **Figure 1-1**.

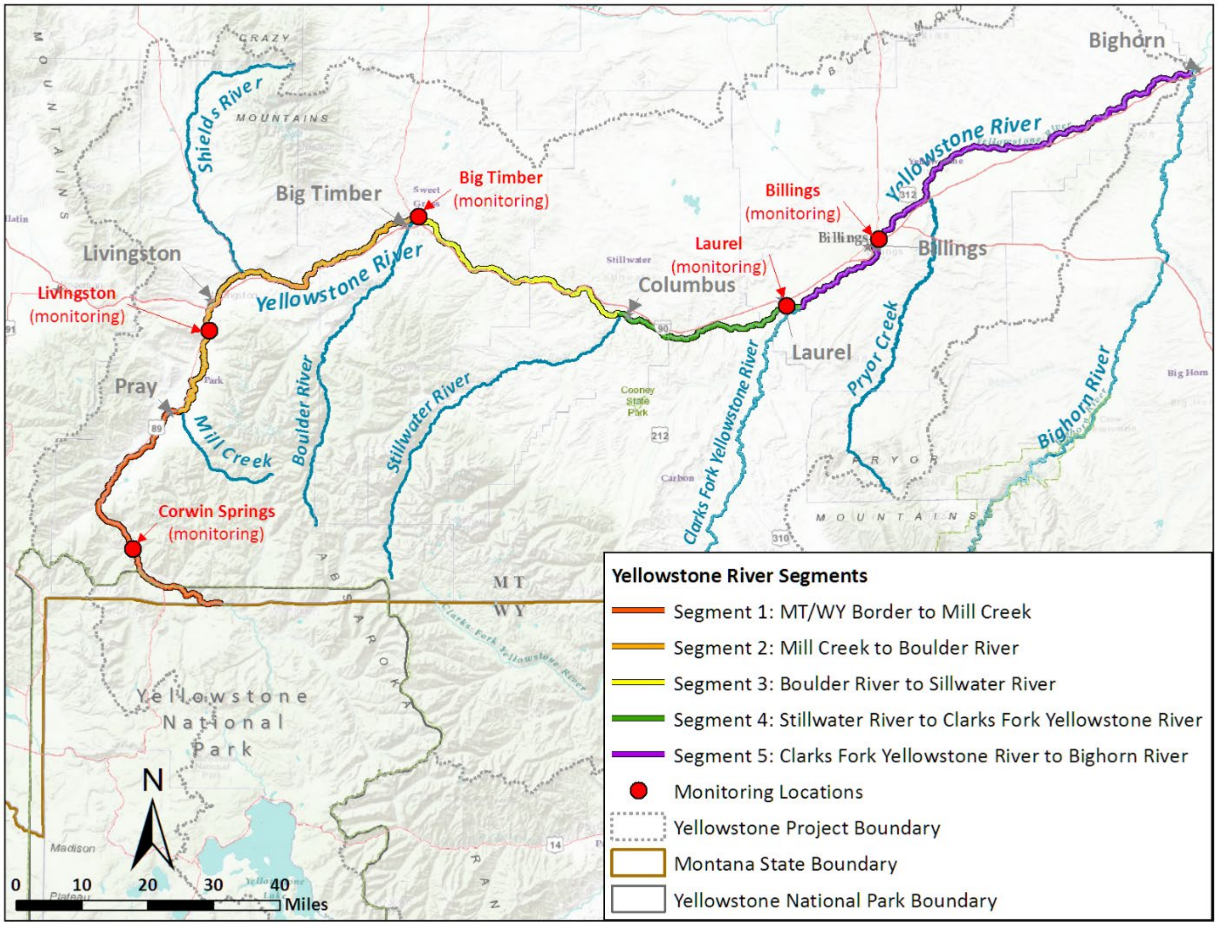

Figure 1-1. Location of Yellowstone River Project Segments and Monitoring Stations

The segments were chosen based on hydrologic divides and a measurable difference in ambient arsenic concentrations (DEQ, 2019). While the entire stretch of the Yellowstone River project reach has at least periodic arsenic concentrations elevated above 10 µg/L due to the geothermal activity in YNP, concentrations decrease in a downstream direction as the Yellowstone River is diluted by flow from tributaries having lower arsenic concentrations (DEQ, 2019).

Hydrograph information evaluated as part of the DON showed that the Yellowstone River has a high flow season from May 1 to July 31 and a low flow season of August 1 to April 30; monitoring data showed that the arsenic concentrations also varies by these high and low flow seasons (DEQ, 2019). The annual and seasonal total arsenic concentrations are depicted in **Figure 1-2.**

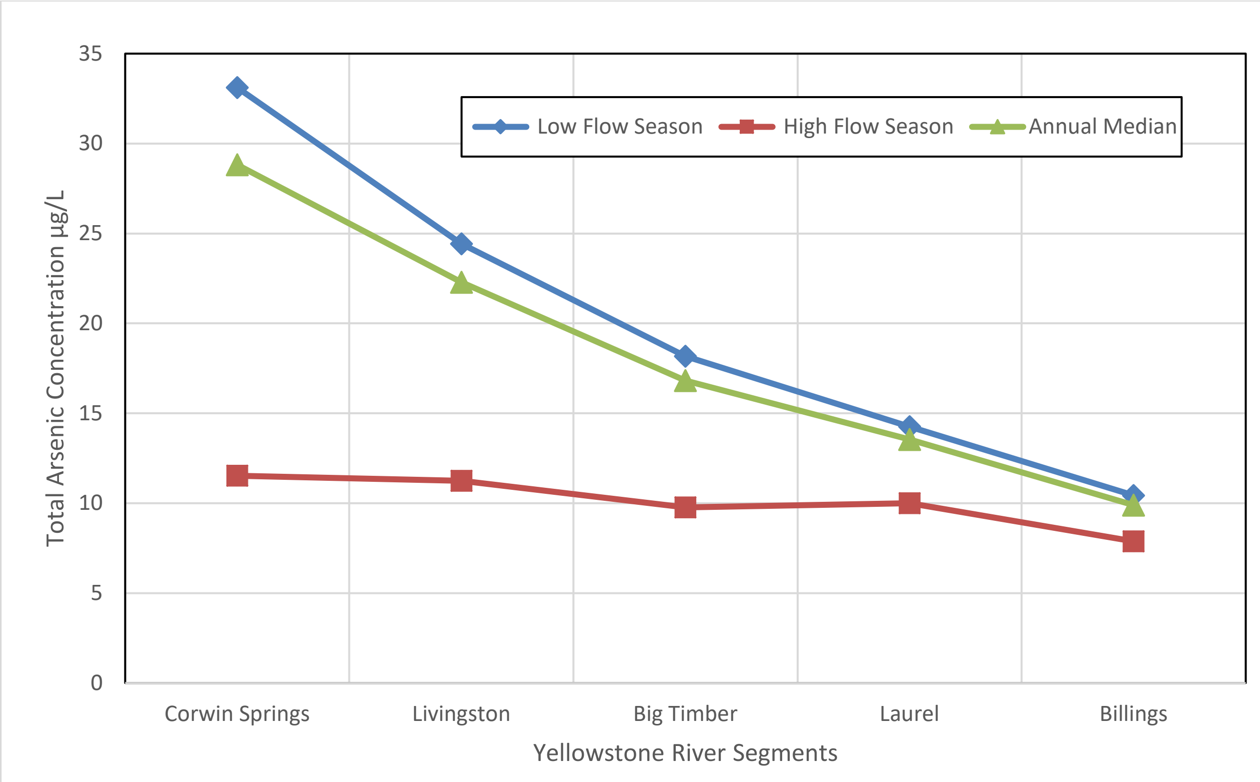

Figure 1-2. Yellowstone River Median Total Arsenic Concentrations

After identifying the hydrologic segments and seasonality, DEQ conducted a modified mass balance in the DON to determine the proportion of the arsenic load in the project reach of the Yellowstone River that was nonanthropogenic; this was carried out for each of the five river segments (DEQ, 2019). The monthly and seasonal nonanthropogenic loads are presented in **Appendix A**. **Table 1-1** describes each segment and provides the percent of the total arsenic load that is nonanthropogenic seasonally (DEQ, 2019).

Table 1-1. Nonanthropogenic Seasonal Arsenic Load Percentages, by Segment

Yellowstone River Segment				Yellowstone River Sampling Location	Proportion of Arsenic Load that is Nonanthropogenic[1]	
#	Beginning	End [2]	Length (miles)		High Flow Season [3]	Low Flow Season [3]
1	Montana/Wyoming Border	Mill Creek near Pray	45	Corwin Springs	99.0%	97.0%
2	Mill Creek	Boulder River at Big Timber	54	Livingston	98.9%	96.9%
3	Boulder River	Stillwater River near Columbus	37	Big Timber	98.9%	96.5%
4	Stillwater River	Clarks Fork of the Yellowstone River at Laurel	27	Laurel	98.9%	95.6%
5	Clarks Fork of the Yellowstone River	Bighorn River at Bighorn	73	Billings	98.7%	95.6%

[1] Based on the median of the LOADEST-modeled daily loads in the DON Appendix C.
[2] Each segment ends immediately before the confluence with the referenced tributary.
[3] High Flow Season for the Yellowstone River was determined to be May – July, and the Low Flow Season was determined to be August - April.

1.3.3 Distribution of Values

Although water quality standards are almost always expressed as a unique concentration, water quality is not a static number. Variability in concentration is often a result of seasonal changes, effects of flow changes, and inter-annual fluctuations. The Yellowstone River at Corwin Springs (**Figure 1-3**) demonstrates the variability of arsenic concentrations over time and the inherent challenge in picking a unique concentration value to represent the “natural” condition of the water body.

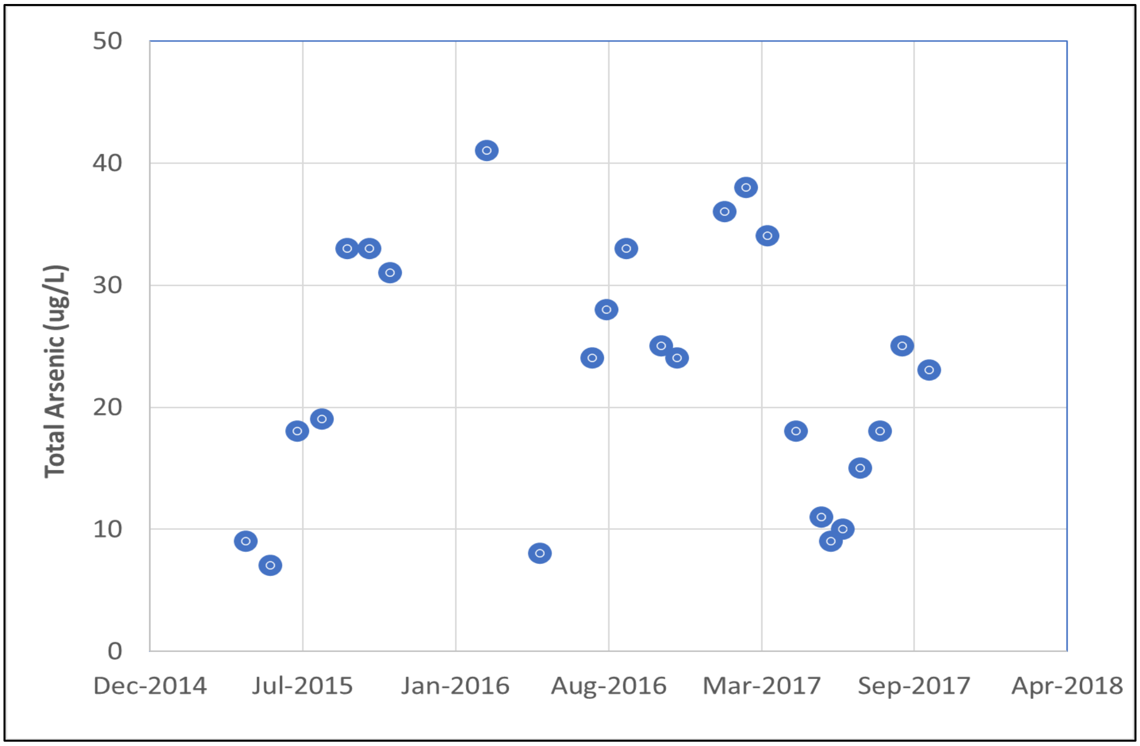

Figure 1-3. Variable Arsenic Concentrations Over Time for the Yellowstone River at Corwin Springs

The purpose of a nonanthropogenic water quality standard is to protect the existing uses of the water body and maintain, as best possible, the long-term nonanthropogenic condition (i.e., the nonanthropogenic distribution of values). While it may not be possible to preserve the exact distribution of values, choosing an appropriate standard within the distribution can ensure that the values necessary to maintain existing uses and conditions are protected.

To achieve that goal, the median of the daily nonanthropogenic concentrations will be used to determine the seasonal standards. By using the median, any change to the frequency distribution of arsenic concentrations induced by anthropogenic means will tend to move the distribution even more to the middle (**Figure 1-4**). **Figure 1-4** also illustrates the potential long-term (and undesirable) shift in arsenic concentrations that might result if the 75^{th} percentile (rather than the median) was established as the standard. It is worth noting that it would take a very large anthropogenic effect to cause such a change. DEQ has identified the segment-specific standards based on the median of the nonanthropogenic concentration distributions; in this manner, the standards will protect the nonanthropogenic arsenic concentrations from shifting up or down.

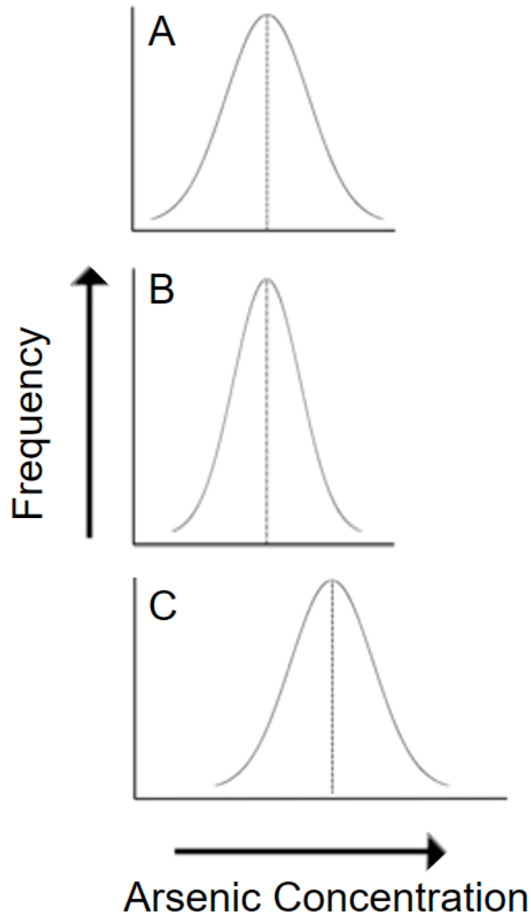

Figure 1-4. Frequency Distribution Scenarios where the Vertical Dotted Lines Represent the Median Arsenic Concentration. A. The existing nonanthropogenic condition, which is normally distributed. B. Scenario in which the standard has been set at the nonanthropogenic median, and the frequency distribution has shifted more towards the center; this could occur if anthropogenic changes were sufficient to alter the existing distribution. C. Scenario in which the 75th percentile was established as the standard and anthropogenic changes were sufficient to move the existing distribution to the right.

2.0 Nonanthropogenic Arsenic Standards Methods

In this document, DEQ will use the findings and conclusions from the DON to develop nonanthropogenic concentration standards for the Yellowstone River segments that have seasonal nonanthropogenic arsenic concentrations equal to or above the human health standard of 10 µg/L. The steps in the DON as well as the NAS (this document) are depicted in **Figure 2-1**. The DON steps were described elsewhere (DEQ, 2019), and will be briefly summarized here, followed by a description of the steps in this document.

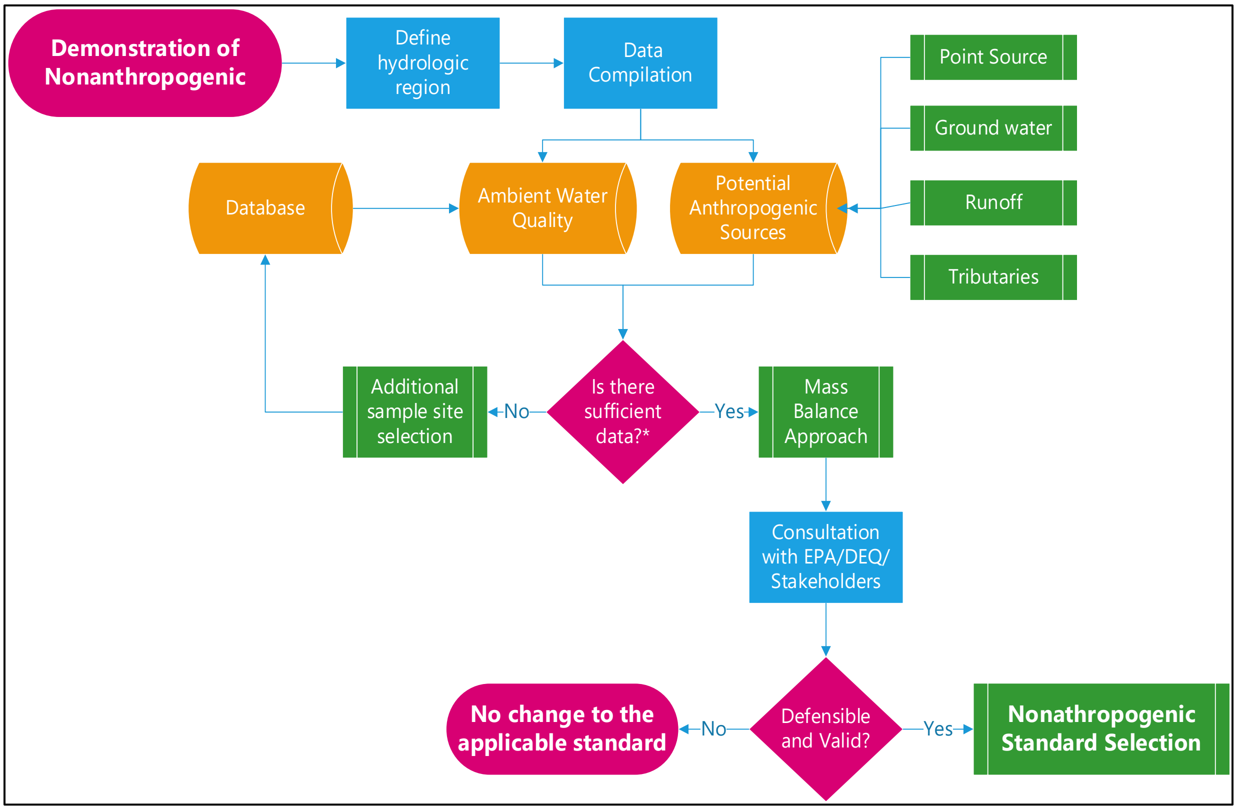

Figure 2-1. Demonstration of Nonanthropogenic Process Detailed in the Arsenic DON (DEQ, 2019)

2.1 Summary of the DON Methods

In the DON (DEQ, 2019), DEQ developed seasonal nonanthropogenic arsenic loads and the percentage of the total arsenic load that was nonanthropogenic; loads were based on the following.

- ***Total arsenic loads.*** DEQ used LOADEST modeling (see the DON Appendix C). The basis of this modeling was sufficient flow (10 years) and measured arsenic concentration (19 to 28 samples) for each of the five segments (DEQ, 2019). For each segment, LOADEST output predicted *daily* loads (i.e., projected total arsenic load in kg/day) and flow rates, for every day between January 1, 2009 and December 31, 2017. From these, DEQ calculated the median monthly, seasonal, and annual total arsenic loads.

- ***Anthropogenic arsenic loads.*** DEQ calculated these loads monthly for each segment using a modified mass-balance (see the DON, Appendix E). From those, DEQ calculated the average daily anthropogenic loads for each month in the DON Appendix C.
- ***Nonanthropogenic arsenic loads.*** DEQ calculated these by subtracting the anthropogenic load from the total arsenic load for each of the five segments. The monthly and annual nonanthropogenic loads and proportion of the total arsenic load that was nonanthropogenic was developed in the DON Appendix E-1 and presented seasonally and annually in the DON Table 4-9 and Table 5-1.

In the DON, DEQ calculated the arsenic loads seasonally based on the following:

- ***Seasonality for flow.*** DEQ identified low- and high-flow regimes based on hydrographs of the Yellowstone River. Evaluations were done per methods in Suplee et al. (2007). For all segments, the high flow runoff period is May 1 to July 31 and the low flow period is August 1 to April 30.
- ***Seasonality for arsenic loads***. DEQ determined that total arsenic concentrations vary under the two flow conditions—the median arsenic concentrations for the high and low flow seasons were significantly different per the Mann-Whitney non-parametric test. Based on this result, DEQ determined that seasonal standards would be appropriate (DEQ, 2019).

The daily nonanthropogenic loads (kg/day), computed by season, were calculated in the DON Appendix C and are presented in **Appendix A** and summarized below in **Table 2-1**. The seasonal loads were calculated as the median of the daily loads within the season.

Table 2-1. Median Nonanthropogenic Arsenic Loads for the Yellowstone River Project Area

Hydrologic Segment	Cumulative Nonanthropogenic Seasonal Loads [1,2] (kg/day)	
	High Flow	Low Flow
1- MT/WY Border to Mill Creek	220	93
2- Mill Creek to Boulder River	249	98
3- Boulder River to Stillwater River	271	88
4- Stillwater River to Clarks Fork Yellowstone River	311	80
5- Clarks Fork Yellowstone River to Bighorn River	317	86

[1] High Flow Season for the Yellowstone River was determined to be May – July, and the Low Flow Season was determined to be August - April.
[2] The seasonal loads were calculated by using the median of the daily loads for the season as developed in the LOADEST output in the DON Appendix C.

2.2 Description of NAS

Water quality standards are expressed as concentrations, not loads. To convert the nonanthropogenic arsenic loads developed in the DON into nonanthropogenic concentration data for the NAS, *Equation 1* is rearranged to solve for concentration in *Equation 1.B*:

Equation 1:	**ML = C x Q x t x cf**	Develop statistically-derived load from concentration and flow data using LOADEST
Equation 1.B:	**C = ML/(Q x t x cf)**	Calculate statistically-derived concentration data

Where,

ML – Nonanthropogenic Mass Arsenic Load (kg/day)
C – Concentration (µg/L)
Q – Flow of water at a point (cubic feet per second, cfs)
t – A period of time (season, month, year; a day in this case)
cf – conversion factor for mass load calculation (variable depending on units of individual terms)

Using *Equation 1.B*, DEQ divided the daily nonanthropogenic load by the daily flow to derive the daily nonanthropogenic arsenic concentration from the LOADEST results. The median of these daily nonanthropogenic arsenic concentrations for each segment were then compiled by month and season. The results are included as **Appendix B** and summarized in **Section 3.0**.

The nonanthropogenic arsenic standards for the Yellowstone River project reach will be based on the calculated median daily nonanthropogenic concentrations for each segment during the high- and low-flow seasons. This approach establishes the water quality standard at a value that is protective locally (i.e., representative of the nonanthropogenic condition), with the median as the best representation of the central tendency of the nonanthropogenic distribution.

3.0 RESULTS

3.1 IDENTIFICATION OF NONANTHROPOGENIC ARSENIC STANDARDS

The median nonanthropogenic arsenic concentrations for each season were calculated based on LOADEST daily nonanthropogenic arsenic loads and the daily flow rates using *Equation 1.B* (**Section 2.2**). The seasonal median concentrations can be used to establish nonanthropogenic arsenic standards(**Table 3-1)**. Monthly values are available in **Appendix B**.

Table 3-1: Yellowstone River Segments and their Median Nonanthropogenic Arsenic Concentrations

Segment	Description	High Flow Season Arsenic Conc.[1] (µg/L)	Low Flow Season Arsenic Conc.[1] (µg/L)
1	MT/WY Border to Mill Creek	11	32
2	Mill Creek to Boulder River	11	24
3	Boulder River to Stillwater River	10	18
4	Stillwater River to Clarks Fork Yellowstone	10	14
5	Clarks Fork Yellowstone to Bighorn River	8	10

[1] High Flow season for the Yellowstone River is May – July, and the Low Flow Season is August - April.

As shown in **Table 3-1** and described below, each segment has different seasonal values:

- Segments 1 and 2 have segment-specific nonanthropogenic arsenic standards for both high and low flow seasons that are above the current human health standard (HHS) of 10 µg/L;
- Segments 3 and 4 have segment-specific nonanthropogenic arsenic standards for the low flow season that are above the HHS. During the high flow season the river's median nonanthropogenic concentrations (and, therefore, the standards) in these segments are equal to the HHS of 10 µg/L; and
- Segment 5 has a high flow season median concentration (8 µg/L) that is lower than the HHS, while its low flow seasonal value is equal to the HHS.

As of this writing, in locations where the river's median nonanthropogenic concentrations are lower than the existing human health standard of 10 µg/L, DEQ is not proposing to change the currently-adopted human-health based standard. This applies to Segment 5 during high flow.

As discussed in **Section 1.3.3**, it is challenging to assign a unique concentration value to represent the condition of a waterbody when its concentrations vary naturally. Case in point, **Figure 3-1** presents the relationship between the proposed seasonal arsenic standards and the median daily nonanthropogenic arsenic concentrations for Segment 1 (Corwin Springs). **Appendix C** presents the same comparison for the remaining segments.

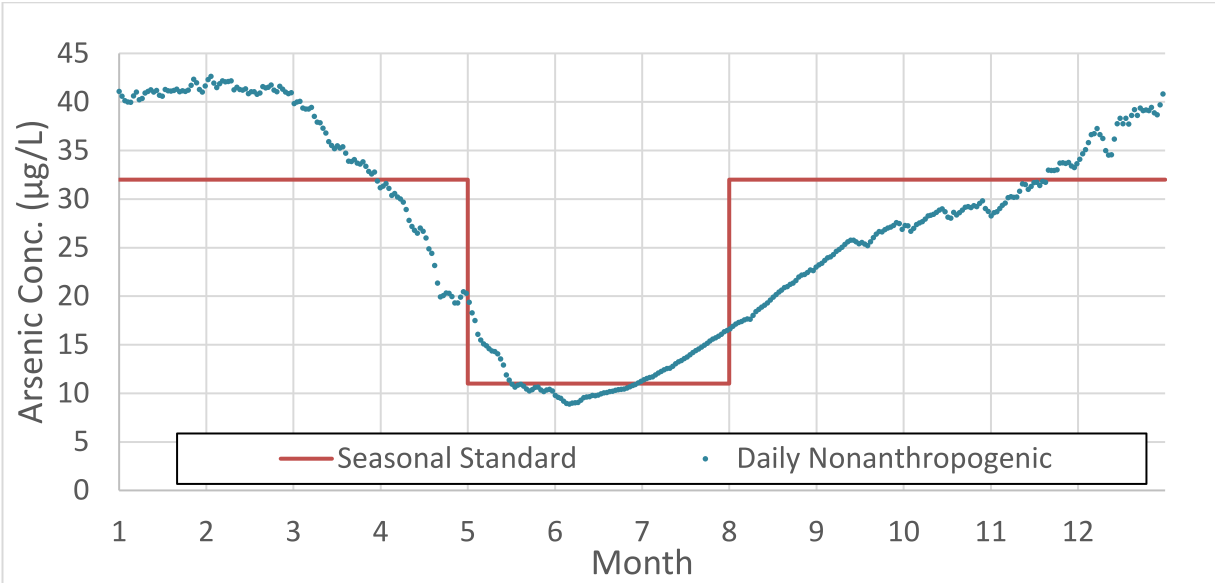

Figure 3-1. Comparison of the Seasonal Standards Against the Median Daily Nonanthropogenic Arsenic Concentrations in Segment 1 – Corwin Springs

The median daily arsenic concentrations for **Figure 3-1** and in **Appendix C** were obtained from 10-year LOADEST modeling conducted in Appendix C of the DON. **Figure 3-1** shows that, even at a monthly scale, natural arsenic concentrations can vary more than 15 µg/L (April) and regularly vary 5 µg/L in a month. However, establishing a water quality standard at a monthly scale rather than a seasonal scale would create water quality assessment and compliance issues, because a relatively large number of samples (>6) must be collected to compute a reasonable sample mean and the samples should, to be statistically valid, meet temporal independence minima intervals (DEQ, 2016b).

Based on these considerations, as well as analysis of the annual hydrographs for the Yellowstone River that showed statistically significant differences in arsenic concentrations between high- and low-flow seasons (DEQ, 2019), DEQ determined that a two-season standard was most appropriate for arsenic on the Yellowstone River.

3.2 Frequency and Duration

A specified frequency and duration must accompany the proposed arsenic standards. Frequency is how often an arsenic value (i.e., the concentrations in **Table 3-1**) can be exceeded without there being a standards violation, while duration refers to the time period over which the data are considered. This section discusses how the frequency (zero exceedances) and duration (seasonal) were developed.

The proposed frequency and duration for the Yellowstone River arsenic standards are: *the median (or average) concentration within each season shall not exceed the standard.* This treats the low flow and high flow seasons as separate periods, and compliance is evaluated for each based only on the measurements collected within the season. The average concentration within each season refers to the arithmetic mean.

Somewhat different approaches will be applied for ambient monitoring *vs.* discharge permit development to meet the frequency and duration statement above. Specifically, ambient river conditions will be evaluated be assessing changes in the median concentration of future datasets (more on this in **Section 4.1**). In contrast, compliance with discharge permits will be based in part on the mean (average) of the effluent concentrations, which aligns with the expression of other water quality standards and conforms with the MPDES permitting program (details will be discussed in **Section 4.2**). Both the mean and median are measures of central tendency, and while the mean and median can be very different in some cases, in this case the mean and median arsenic concentrations in the Yellowstone River are very similar (**Table 3-2**). This indicates that the river's arsenic concentration data are normally distributed (i.e., they form a normal bell-shaped curve). Therefore, it is reasonable to use one measure to establish the standard and one measure to enforce the standard.

Table 3-2. Yellowstone River Segments Comparison of 2008-2018 Total Arsenic Annual Median and Mean (Average) Values (in µg/L) from LOADEST [1]

Statistic	Corwin Springs	Livingston	Big Timber	Laurel	Billings
Annual Median	28.8	22.3	16.8	13.5	9.9
Annual Average	27.7	21.2	16.2	13.3	10.0
1 Annual Median and Average of Daily Values from Appendix C (electronic) in DON.					

3.3 Highest Attainable Use

A critical step in applying nonanthropogenic arsenic standards to the Yellowstone River is determining the highest attainable use(s) under the nonanthropogenic condition[1]. As noted in **Section 1.3.1**, arsenic is a human-health carcinogen and the current standard (10 µg/L) is intended to protect water used for drinking. The Yellowstone River's beneficial uses which pertain to drinking water are "drinking, culinary, and food processing purposes, after conventional treatment" (ARM 17.30.611; ARM 17.30.623 through 625).

For the Yellowstone River, adoption of nonanthropogenic arsenic standards will not require changes to the river's designated beneficial uses. The highest anticipated instream arsenic concentration estimated using the LOADEST modeling is approximately 60 µg/L in segment 1 (Montana/Wyoming border to Mill Creek), and segments 2 through 4 have lower concentrations. According to DEQ drinking water engineers (Denver Fraser, personal communication, 9/23/2019), arsenic at an intake concentration of 60 µg/L or less can be treated to below the arsenic human health standard of 10 µg/L (Circular DEQ-7) using conventional treatment as defined in ARM 17.30.602(5). Since current law assumes conventional treatment is required before the river's water can be rendered drinkable, there is no need for DEQ to make changes to the river's beneficial drinking water use.

[1] Note: state and federal regulations (75-5-301 and 302, MCA; ARM 17.30.606 and 621 through 629; and 40 CFR 131.10), and federal guidance on use designation are available from DEQ and the Environmental Protection Agency.

4.0 Implementation

Identification of the nonanthropogenic arsenic standards (**Section 3.1**) does not complete the standards development process. To implement the standards, consideration must be given to how discharge permits will be written, how ambient river concentrations will be assessed going forward, how DEQ will ensure continual protection of downstream water quality standards, etc. These subjects are covered in this section. Included are:

1. A method for assessing changes in ambient arsenic concentrations; and

2. Effluent limit calculations, including determination of reasonable potential, consideration of non-degradation, and downstream use protection.

4.1 Method for Assessing Ambient River Conditions

In the long run, water quality and beneficial use assessments determine if water quality continues to meet the level of nonanthropogenic water quality originally characterized by the nonanthropogenic standards. Because the standards are based on the nonanthropogenic median and nearly all of what we measure in the river is nonanthropogenic, it would be expected that, over the long term, half of the future years assessed could exceed the standard and half will not, based on natural variability alone. To account for this natural variability, the following assessment method has been developed.

4.1.1. Wilson Interval Method (Confidence Interval Method)

A statistical approach, based on confidence intervals, will be used to determine whether a Yellowstone River dataset assessed in the future can be considered significantly different from the nonanthropogenic condition that was used to define the standard. A confidence interval is most easily understood as the region around a value (the median of an assessed dataset in this case) within which the true value is likely to be located (CDPHE, 2013). The width of the confidence interval, and therefore the range of values it spans, is determined in part by the desired level of confidence. There are several ways to calculate a confidence interval, but the method used here is based on the Wilson Interval – described and used in water quality assessments by the Colorado Department of Public Health and Environment (CDPHE, 2013). A major advantage of the Wilson Interval is that a confidence interval of a chosen level (say, 90% confidence) can be established around any percentile in the test dataset that is deemed important (the 30th, 50th, 64th, etc.; whatever is meaningful). Because the arsenic standards are being established at the nonanthropogenic median (50th percentile), the 50th percentile is the meaningful percentile we want to evaluate in future assessments of the ambient river condition[2].

Figure 4-1 below shows a non-exceedence scenario and an exceedance scenario using the Wilson Interval test. Whenever the concentration corresponding to the Wilson lower confidence limit (LCL) of a

[2] DEQ considered using the One Sample T-test for the Mean, which could also compare change in a measure of central tendency (future *average* arsenic concentration) to a nonanthropogenic standard. However, that test is greatly influenced by outliers (CEPA, 2004). Outliers are likely to be part of the small datasets used for assessing the river going forward.

future dataset is higher than the nonanthropogenic standard concentration, an exceedance has occurred (**Figure 4-1B**).

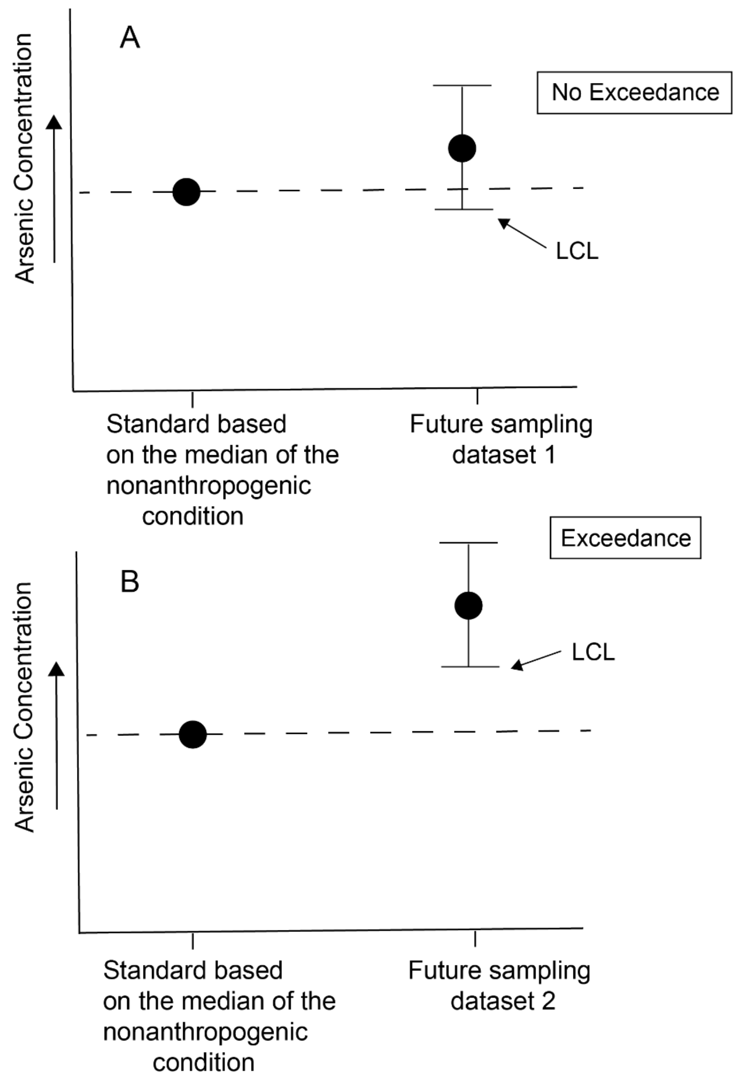

Figure 4-1. Conceptual Representation of the Wilson Interval Method to Determine Compliance with an Arsenic Standard. A. Scenario where no exceedance has occurred because the standard falls in the range between the future dataset's median (black dot) and its lower confidence limit (LCL). B. Scenario where an exceedance has occurred, because the standard is beyond the LCL of the future sampling dataset.

Both sample size and confidence level (CL) of the confidence interval influence the Wilson Interval's sensitivity to detect change. DEQ sought to identify the minimum number of samples which could provide a reasonable degree of sensitivity. To do this, we carried out an analysis on low-flow (August-April) arsenic concentration data DEQ collected from 2015-2017 along the Yellowstone River ("current conditions datasets"). We established two *a priori* criteria to evaluate sensitivity:

1. For the current conditions datasets, the LCL of the Wilson Interval should usually fall between the median and the lower quartile (25th percentile) of the dataset.

2. For the current conditions datasets, the false positive rate (i.e., concluding the river exceeds the standard when in truth it does not) should be less than ≤10%.

Since the arsenic standards are being established very close to current ambient conditions, there was no simple way to evaluate false *negative* rates because, by definition, there are no sites on the river that exceed the standards (and these would be necessary to inform a false negative analysis). Instead, we established criterion 1 above; if the LCL of a particular sample size/CL combination fell below the interquartile range (say, near the 10th percentile) of current, representative data, then we would conclude that that combination would not be very sensitive to change.

We assessed sample size/CL combinations using the two criteria by bootstrapping the 2015-2017 arsenic datasets[3]. Bootstrapping is a statistical re-sampling technique that randomly samples from an original dataset with replacement. It can be used for estimating summary statistics for a population from small data samples (Efron, 1993). In each segment, for each sample size of interest and for each CL of interest, the raw data were bootstrapped 1000 times (Manley, 1997); e.g., in Segment 1 for sample size five we randomly generated 1000 sampling events of n=5. The mean response of these 1000 events was taken to be representative of that sample size. We bootstrapped samples sizes ranging from 5 to 28 and computed corresponding Wilson LCLs for the 85th, 90th, 95th, and 99th confidence levels. False positive rates were computed by determining the number of sampling events (of 1000) where the LCL exceeded the applicable low-flow standards in **Table 3-1**.

Some general patterns emerged. At low sample sizes (usually less than about 10) and/or high CLs, criterion 1 was not met very often. Another generality was that false positive rates decreased with increasing CLs. As sample size increased to 14-16 and higher, and at CLs of 85-90%, most Wilson LCLs fell within the interquartile range while still maintaining low false positive rates (**Table 4-1**). Another general pattern was that Segments 1-3 were sufficiently similar that a single sample size/CL combination could be recommended for all three, whereas Segment 4 had a much higher false positive rate and warranted a different sample size/CL. In Segments 1-3, the river's median arsenic concentrations would have to increase about 20% (~6 µg/L) for DEQ to conclude that an exceedance had occurred (**Table 4-1**). In Segment 4, the interquartile range of the arsenic concentrations is much narrower, which means that detectable change is more sensitive; a change in the median of only 10% (about 1 µg/L) is necessary for DEQ to conclude that an exceedance had occurred. (Segment 4's narrower interquartile range is also the cause of its higher false positive rate.)

Specific recommendations for sampling the segments are provided in the next section.

[3] DEQ will provide the bootstrap analytical spreadsheet, which includes the test datasets, upon request. Please contact staff of the Water Quality Standards & Modeling Section.

Table 4-1. Sensitivity of the Wilson Interval in Segment 1 for Selected Sample Sizes and Confidence Levels.

Confidence Level (%)	Sample Size											
	14				15				16			
	% of sampling events where LCL fell between dataset median and lower quartile	False Positives (%)	Change needed to trigger exceedence (µg/L)	Change needed to trigger exceedence (%)	% of sampling events where LCL fell between dataset median and lower quartile	False Positives (%)	Change needed to trigger exceedence (µg/L)	Change needed to trigger exceedence (%)	% of sampling events where LCL fell between dataset median and lower quartile	False Positives (%)	Change needed to trigger exceedence (µg/L)	Change needed to trigger exceedence (%)
85	70	3.2	5.7	17.9	67	2.6	5.6	17.6	65	2	5.5	17.1
90	54	2	6.5	20.3	66	1.3	6.3	19.7	65	1.4	6.2	19.4
95	53	0.7	7.6	23.7	51	0.9	7.4	23.0	46	0.3	7.2	22.5
99	31	0	9.4	29.5	28	0.1	9.2	28.7	25	0.2	8.9	27.9

4.1.2 Seasonal Data Needs, Spatial Considerations, Sample Sizes and Confidence Levels

DEQ's Monitoring & Assessment Section will collect (or compile) data specific to both high flow and low flow seasons and assess each segment and season on its own merits. Low and high flow seasons are each addressed below.

- ***Low Flow.*** In Segments 1-3, from 14-16 samples should be collected in each segment over two calendar years (14-16 samples total, not per year). A minimum sampling interval between sampling events of two weeks should be maintained to assure temporal sample independence (DEQ, 2016b). Each dataset should then be assessed using the Wilson Interval specific to the 50th percentile (p-hat = 0.5) with a confidence level set at 90%. In segment 4, we recommend a minimum of 15 samples with the same two-week timespan between events as above, collected over two calendar years. The dataset should then be assessed using the Wilson Interval specific to the 50th percentile (p-hat = 0.5) with a confidence level set at 95%. Based on the bootstrap analyses, this should keep the false positive rate for Segment 4 under 10%.

- ***High Flow.*** Three important factors were given consideration for the High Flow assessment: (1) the bootstrap analysis was based only on low-flow data; (2) the high-flow sampling season is much shorter (three months/year); and (3) the river's volume is up to an order of magnitude higher than during low flow. Because of the latter, there is a much lower likelihood an anthropogenically-caused arsenic exceedance can occur during high flow. Given these considerations, we recommend for Segments 1-4 that 12 samples should be collected in each segment over two calendar years (12 samples total, not per year). A minimum sampling interval between sampling events of two weeks should be maintained to assure temporal sample independence (DEQ, 2016b). Each dataset should then be assessed using the Wilson Interval specific to the 50th percentile (p-hat = 0.5) with a confidence level set at 90%.

Arsenic data may be collected by DEQ or come from sources other than DEQ. If a dataset comprises multiple sampling sites within a segment, the different sites' data collected on the same day (or very close to the same day) will be collated and reduced to a single average value to represent that day (i.e., they are NOT considered spatially independent).

4.1.3 Exceedance Frequency

The Monitoring & Assessment Section should include (or compile, if there are extant data) a dataset which meets the minimums for each season and segment as described in **Section 4.1.2**. Based on the outcomes from the Wilson Interval calculations:

1. Zero (0) exceedances = full compliance
2. One (1) exceedance = non-compliance

An exceedance is not a single river sample above the standard; it is a statistically significant change in the river's median arsenic concentration, represented by the Wilson Interval. The regularity at which DEQ's Monitoring & Assessment Section carries out assessments on the river is at the discretion of the section manager and/or included in the section's long-term monitoring strategy.

Start and end points of Yellowstone River *assessment* units—which are the base units for assessment purposes—do not align with the segment start and end points in this document (assessment units can be found in DEQ's Clean Water Act Information Center, on DEQ's website). For purposes of assessment, any segment in this document (1 through 5) and season (high or low flow) found to be out of compliance within an assessment unit would result in the entire Yellowstone River assessment unit being considered out of compliance.

If the water body is determined non-compliant for arsenic, the water body will either be referred for a TMDL or for redevelopment of the standard depending on the suspected cause of the exceedance. If there is no evidence that the standard has been exceeded due to anthropogenic causes, it may be necessary to establish a new natural standard; that scenario leads back to the Water Quality Standards & Modeling Section (WQSM) in **Figure 4-2**.

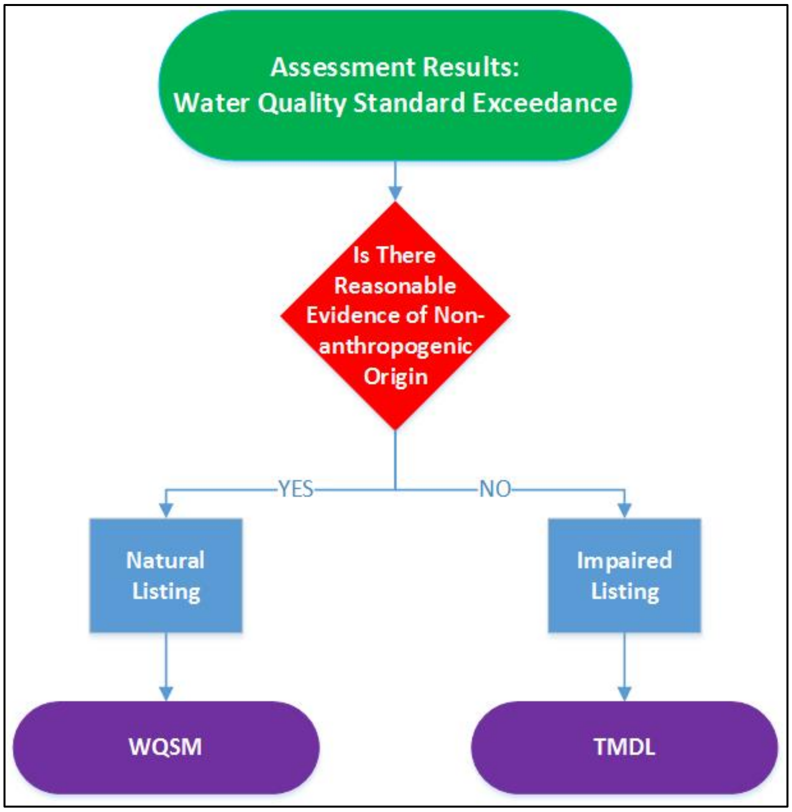

Figure 4-2. Decision Chart for Actions after Finding an Arsenic Standard Exceedance

4.2 Wastewater Discharge Permitting

DEQ issues Montana Pollutant Discharge Elimination System (MPDES) permits to owners/operators of point source discharges who discharge pollutants to state surface waters. Components of the MPDES permit development include effluent and receiving water characterization, reasonable potential analysis, and calculation of effluent limits. The discharge cannot have a reasonable potential to cause or contribute to an exceedance of a water quality standard. For new or increased sources, the review includes a nonsignificance determination and consideration of the protection of water quality of downstream water bodies. These components are further discussed in the sections below.

4.2.1 Reasonable Potential Analysis

Reasonable potential analysis determines whether a discharge has reasonable potential to cause or contribute to an exceedance of a water quality standard. All estimates assume the discharge and receiving water are at critical conditions. These conservative assumptions are used to ensure an impact is not projected to occur (EPA, 1991).

In determining reasonable potential, DEQ will consider controls on point sources and the variability of the pollutant parameter in the effluent. DEQ may also consider any dilution available from the receiving water (if the source has an approved mixing zone); however, because the arsenic standards are set at the nonanthropogenic condition of the Yellowstone River, assimilative capacity does not exist, and no dilution or mixing zone will be granted.

With standards based on nonanthropogenic conditions, if a proposed discharge has the potential to cause or contribute to the exceedance of the nonanthropogenic standard for the receiving water body, reasonable potential exists and necessitates effluent limits (see **Section 4.2.3**).

4.2.2 Application of Nondegradation to Water Quality Standards

For waterbodies where a nonanthropogenic standard is adopted, effluent limits for a new or increased source must be derived to protect the nonanthropogenic concentration of that parameter in the receiving water. Effluent limits based on a nonanthropogenic standard will also ensure existing and anticipated uses are maintained and protected pursuant to ARM 17.30.705(2)(a). Additional nondegradation protection is not applicable to new or increased sources because the water quality standard is established at the nonanthropogenic instream concentration, and because the department may not apply a standard to a waterbody that is more stringent than the nonanthropogenic condition of the waterbody.

4.2.3 Effluent Limit Calculations

The seasonal nonanthropogenic standards developed in this document will be incorporated as the wasteload allocations (WLAs) when calculating permit limits. Since the standards are established at existing nonanthropogenic arsenic conditions there is no assimilative capacity; therefore, DEQ will not grant a mixing zone and will not provide dilution when it develops permit limits.

In the case of arsenic, since the seasonal nonanthropogenic human health standard is more conservative than the aquatic life standards, the WLA will be equivalent to the average monthly limit. Effectively, this means the seasonal nonanthropogenic arsenic standard will become the end-of-pipe average monthly limit for the applicable months.

Although the median condition of the river was used to define the nonanthropogenic standards (**Section 3.1**), the nonanthropogenic standards will be applied as the average (arithmetic mean) effluent limit to keep in line with other permitting requirements, simplify calculations, and provide more protection from extreme variations in point source discharges (if DEQ used the median for this purpose, it would be less sensitive to high outliers).

In addition to the average monthly limit, DEQ calculates a maximum daily limit for all toxics. The maximum daily limit will be calculated using methods suggested in the Technical Support Document for Water Quality-based Toxics Control (EPA, 1991) or another method approved by DEQ and accepted by the Environmental Protection Agency (EPA).

4.2.4 Protection of Downstream Water Quality

Prior to DEQ issuing or renewing a MPDES permit implementing the nonanthropogenic arsenic standards, DEQ will ensure downstream water quality is protected.

5.0 REFERENCES

CDPHE (Colorado Department of Public Health and Environment), 2013, Assessing Attainment of Ambient-Based, Water Quality Standards, Colorado Department of Public Health and Environment, From: WQCD 2016 303(d) Listing Methodology Administrative Action Hearing Draft for Commission Review - March 2015 published Jan 8, 2015c.

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

DEQ, 2019, Demonstration of Nonanthropogenic Arsenic Levels: Yellowstone River, Montana. Helena, MT: Montana Dept. of Environmental Quality.

-----, 2017a, Montana Final 2017 Water Quality Integrated Report, WQPBIMTSTR-011f, Montana's Clean Water Act Information Center, Montana Department of Environmental Quality, http://deq.mt.gov/Water/WQPB/cwaic/reports, Accessed 2/2/2017

-----, 2017b, Sampling and Analysis Plan, Natural Arsenic Concentrations: Yellowstone and Upper Missouri River Basin, WQPBWQSMSAP-01, Water Quality Standards Section, Water Quality Planning Bureau, Montana Department of Environmental Quality.

-----, 2017c, Statewide Monitoring Data 1995-2016, Unpublished work.

-----, 2016a, Sampling and Analysis Plan, Natural Arsenic Concentrations: Yellowstone and Upper Missouri River Basin, WQPBWQSSAP-15, Water Quality Standards Section, Water Quality Planning Bureau, Montana Dept. of Environmental Quality

-----, 2016b, Assessment Methodology for Determining Wadeable Stream Impairment Due to Excess Nitrogen and Phosphorus Levels. Helena, MT: Montana Dept. of Environmental Quality.

----, 2015a, Quality Assurance Project Plan: Natural Arsenic Concentrations: Yellowstone and Upper Missouri River Basin, WQPBQAP-15, Water Quality Standards Section, Water Quality Planning Bureau, Montana Dept. of Environmental Quality

-----, 2015b, Sampling and Analysis Plan, Natural Arsenic Concentrations: Yellowstone and Upper Missouri River Basin, WQPBWQSSAP-15, Water Quality Standards Section, Water Quality Planning Bureau, Montana Department of Environmental Quality

-----, 2012, DEQ-7 Montana Numeric Water Quality Standards, Water Quality Standards Section, Water Quality Planning Bureau, Montana Dept. of Environmental Quality, https://deq.mt.gov/Portals/112/Water/WQPB/Standards/PDF/DEQ7/FinalApprovedDEQ7.pdf

Efron, B., 1993, An Introduction to the Bootstrap. Chapman and Hall/CRC.

EPA, 1991, Technical Support Document for Water Quality-Based Toxics Control, Environmental

Protection Agency, Document 5052-90-001, http://water.epa.gov/scitech/datait/models/upload/2002_10_25_npdes_pubs_owm0264.pdf

Manly, B. F. J., 1997, Randomization, Bootstrap and Monte Carlo Methods in Biology, Second Edition, Chapman and Hall/CRC

Suplee, Michael W., Arun Varghese, and Joshua Cleland, 2007. Developing Nutrient Criteria for Streams: An Evaluation of the Frequency Distribution Method. Journal of the American Water Resources Association (JAWRA) 43(2):453-472.

USGS, 2015, Streamflow Characteristics Based on Data through Water Year 2009 for Selected Streamflow-Gaging Stations in or near Montana, U.S. Geological Survey Scientific Investigations Report 2015-5019.

YNP, 2015, Yellowstone National Park Facts, Yellowstone National Park, http://www.nps.gov/yell/naturescience/geology.htm. Accessed 2/1/2015

APPENDICES

Appendix A: Nonanthropogenic Median Daily Loads

Median Daily Nonanthropogenic Arsenic Load (kg/day)					
	Corwin Springs	Livingston	Big Timber	Laurel	Billings
October	98.1	103.5	92.7	88.1	98.3
November	92.4	96.8	87.3	80.7	91.7
December	84.9	88.7	78.3	70.1	77.1
January	79.4	83.6	72.9	64.4	73.8
February	79.1	83.6	72.6	63.0	73.0
March	84.1	88.1	78.4	67.0	74.6
April	103.8	107.6	102.0	96.3	103.9
May	207.3	232.1	249.7	264.8	268.6
June	285.0	331.0	386.9	491.0	525.8
July	185.2	199.5	198.4	224.8	204.6
August	132.0	129.2	115.5	114.5	102.3
September	105.5	106.7	94.3	87.7	85.9
Low Flow Season	**91**	**96**	**86**	**78**	**84**
High Flow Season	**220**	**249**	**271**	**311**	**317**

Appendix B: Yellowstone River Median Nonanthropogenic Arsenic Concentration Calculated by LOADest (2008 - 2018)

Median Daily Nonanthropgenic Arsenic Concentration (µg/L)					
	Corwin Springs	Livingston	Big Timber	Laurel	Billings
January	41.4	28.5	20.6	15.1	10.9
February	41.8	28.5	20.7	15.3	11.0
March	37.1	26.1	18.9	14.5	10.3
April	26.0	20.3	14.9	12.2	8.9
May	12.0	11.5	9.9	9.9	7.6
June	9.4	9.8	9.0	9.8	7.8
July	13.3	12.6	10.6	10.2	7.9
August	18.9	17.2	13.9	11.7	9.2
September	25.4	20.8	16.1	12.9	9.4
October	28.5	21.6	16.4	12.9	9.3
November	31.5	23.3	17.3	13.4	9.6
December	36.5	26.1	19.1	14.4	10.5
Low Flow Season	**32**	**24**	**18**	**14**	**10**
High Flow Season	**11**	**11**	**10**	**10**	**8**

Appendix C - Nonanthoropogenic Arsenic Standards vs Median Daily Arsenic Concentrations