RMP
REGIONAL MONITORING PROGRAM FOR WATER QUALITY IN SAN FRANCISCO BAY

sfei.org/rmp

North Bay Selenium Monitoring Design

Prepared by

Tom Grieb, Sujoy Roy, and John Rath: Tetra Tech Inc.

Robin Stewart[1]: US Geological Survey

Jennifer Sun and Jay A. Davis: San Francisco Estuary Institute

[1] Contributed to Section 5

Contribution #921 December 2018

SAN FRANCISCO ESTUARY INSTITUTE • 4911 CENTRAL AVE., RICHMOND, CA • WWW.SFEI.ORG

This work was funded as a result of settlement of

San Francisco Bay Water Board enforcement actions

Suggested citation:

Grieb, T., S. Roy, J. Rath, R. Stewart, J. Sun, and J.A. Davis. 2018. North Bay Selenium Monitoring Design. San Francisco Estuary Institute, Richmond, CA. SFEI Contribution #921.

Acknowledgements

This report was prepared under the oversight of the RMP Selenium Workgroup. In particular, comments from Harry Ohlendorf helped to significantly improve the report.

Table of Contents

List of Figures

List of Tables

Executive Summary

The San Francisco Bay Regional Water Quality Control Board developed a Total Maximum Daily Load (TMDL) and implementation plan for selenium in North San Francisco Bay (NSFB). The TMDL is based on attainment of water column and fish tissue target concentrations protective of human health, aquatic life, and wildlife, and was approved by the U.S. Environmental Protection Agency (USEPA) in August 2016. The USEPA also published proposed aquatic life and aquatic-dependent wildlife criteria for selenium in the Bay and Delta in 2016. Subsequently, the San Francisco Water Board reached out to the Selenium Workgroup of the Regional Monitoring Program for Water Quality in San Francisco Bay (RMP) to consider the development and implementation of a robust selenium monitoring plan for the North Bay. A primary goal is to identify leading indicators of change to allow prompt management response to signs of increasing risk of impairment.

This document describes the evaluation of three key selenium indicators of water quality conditions in the North Bay: aqueous concentrations, bivalve tissue concentrations, and fish tissue concentrations. These indicators have been monitored locally at various levels of intensity since the late 1980s. Collectively the historical data provide a robust array of environmental knowledge. This report provides a summary of this information and the factors that affect the observed variability in individual indicator values. Next, the use of these data in statistical analyses to evaluate alternative monitoring strategies capable of detecting specified levels of change in each indicator of selenium change in NSFB is described. These analyses included both an evaluation of the appropriateness of different statistical approaches and the performance of the proposed test methods.

One of the main contributions of this design effort is the evaluation of expected monitoring program performance. Each of the indicators was evaluated in terms of the statistical power (defined as the probability that the dataset will be sufficiently sensitive to detect a change or trend of a specified magnitude) that will be achieved. Overall, the findings show that the implementation and continuation of long-term monitoring programs are required to identify changes from established baselines and to distinguish deviations from the effects of natural variability. For the water indicator, the evaluation of the identified monitoring designs suggested that greater than an 80% likelihood of detecting a 20% change can feasibly be achieved with detection times within a 10-year sampling period. The recommended bivalve monitoring approach was based on detecting deviations from the historical distribution of selenium tissue measurements that is built upon 22 years of monitoring. The recommended monitoring program is shown to achieve a 90% likelihood of detecting exceedance of the 97.5th quantile of the distribution of the existing data with 6 samples collected annually. For fish tissue, the analyses suggested that the recommended design (biennial analysis of 60 samples) would achieve a greater than 80% likelihood of detecting a 2% annual increase within a 20-year sampling period.

1 Introduction and Background Information

The North Bay appeared on the 303(d) list under the Clean Water Act because selenium was identified as causing an impairment of the Bay's existing beneficial uses, including estuarine habitat, preservation of rare and endangered species, and sport fishing (Baginska, 2015). In April 2014, the Regional Monitoring Program for Water Quality in San Francisco Bay (RMP) formed a Selenium Workgroup to evaluate information needs and to provide oversight for studies to support development and implementation of selenium regulations for the Bay. The San Francisco Bay Regional Water Quality Control Board developed a Total Maximum Daily Load (TMDL) and implementation plan for selenium in North San Francisco Bay (NSFB) in 2015 (Baginska, 2015). The TMDL is based on attainment of water column and fish tissue target concentrations protective of human health, aquatic life, and wildlife, and was approved by the U.S. Environmental Protection Agency (USEPA) in August 2016. The TMDL established a total dissolved water-column selenium target of 0.5 µg/L and a white sturgeon muscle tissue target of 11.3 µg/g dry weight (dw) (Baginska, 2015). Water-column concentrations are generally well below the 0.5 µg/L target, but fish concentrations are closer to the 11.3 µg/g target and sometimes exceed it. In June 2016, the USEPA published proposed aquatic life and aquatic-dependent wildlife criteria for selenium in the Bay and Delta (USEPA, 2016). The proposal includes criteria for fish tissue (11.3 µg/g dw in muscle and 8.5 µg/g dw in whole body), clam tissue (15 µg/g dw), and water (0.2 µg/L dissolved and 1 µg/g dw particulates).

The Sacramento and San Joaquin Delta is a major source of selenium to the Bay (Baginska, 2015), and there is a concern that proposed future changes in the Delta, by way of new infrastructure (such as the large-scale WaterFix project (California Waterfix, 2018)) or through operations of the export facilities may adversely affect selenium levels in the Bay. Following up on discussions surrounding the North Bay TMDL, the San Francisco Water Board asked the Selenium Workgroup to consider the development and implementation of a robust monitoring plan for the North Bay. A primary goal is to identify leading indicators of change to allow prompt management response to signs of increasing risk of impairment. Of concern are the possible impacts of changes in hydrology in the Delta or changes in selenium loads to Bay-Delta tributaries in the Central Valley.

The Selenium Workgroup convened a technical workshop to address monitoring strategies in July 2016. Specific management questions were articulated, and follow-up discussions provided the basis for identifying the following initial components of the monitoring plan for the North Bay:

1. *Are the beneficial uses of North San Francisco Bay impaired by selenium?*
   This question suggests the need for trend monitoring of matrices identified by regulatory targets:
   - sturgeon muscle and water; and
   - other matrices that are included in draft and final USEPA site-specific criteria.

2. *Are changes in selenium concentrations occurring that warrant changes in management actions?*
   The ability to provide the information to address this question is a key consideration in designing a monitoring program. The Workgroup identified the need to explore an

early warning system for selenium exposures in the Bay ecosystem that would provide multiple lines of evidence of potential impairment. Key design questions that were identified include:

- What would we expect to be the leading indicators of change?
- What would be the yellow flags that an increase is occurring?
- If fish tissue is not a leading indicator of change, what is the expected time lag between changes in selenium in water/particulates/prey and fish tissues?
- What combination of metrics would be most effective in detecting change?
- Would sampling a combination of metrics alter the monitoring scheme (e.g., frequency, location) for fish tissue sampling?

3. *Will proposed changes in water flows and/or selenium loads in the Bay or upstream cause impairment in the North Bay?*
The Workgroup noted that in designing the monitoring program it is essential to evaluate the potential magnitude of the effects of changes that may cause selenium concentrations to change in the NSFB and Delta. The identified changes are: water-column concentration changes in the inflow from the San Joaquin River basin, changes in refinery inputs, changes in stormwater and tributary loads from the Bay margin, and changes in overall Central Valley hydrological conditions, such as the extreme wet and dry periods that occurred between 2012 and 2017. Other factors, such as nutrient concentrations and algae levels, may also play a role, especially on the concentrations of selenium on and in particulates. Over the longer term, selenium changes may occur due to modification of Delta flows and the mix of riverine sources because of the implementation of the WaterFix project by the state of California.

   The Workgroup concluded that there is sufficient concern regarding the potential changes in selenium loads to warrant the consideration of a sustained monitoring effort in the NSFB and Delta to support the long-term implementation goals of the TMDL.

2 Evaluation of the Monitoring Strategies for Each Indicator

In addition to the articulation of the key management questions, the Selenium Workshop included presentations on the historical information available for the three key selenium indicators of existing conditions (i.e., aqueous concentrations, clam tissue concentrations and fish tissue concentrations). These presentations provided summaries of how well we understand the factors (e.g., seasonal effects, differences among sample locations and water-year type) that affect variability in measurements of the three indicators. Based on understanding of the variability in historical measurements, some initial estimates were presented on the anticipated performance of monitoring programs (i.e., what is the ability to detect change, or what is the minimum difference in selenium measurements over time and/or between sampling locations that can be detected for specified levels of sampling effort).

The Selenium Workgroup designated subgroups that worked independently to develop early-warning monitoring programs for aqueous concentrations, clam tissue concentrations and fish tissue concentrations. At the beginning of the process, the Workgroup also specified a common format for development of these monitoring programs. In addition to the development of an overall statement of the objectives of the selenium monitoring program (i.e., the development of an early warning detection system and the detection of long-term trends), the following six elements were identified to provide a common framework for the evaluation of monitoring the three indicators:

A. **Overview of the Monitoring Element**
The overview provides a review and summary of the historical data. This includes a synopsis of what we have learned, and how the existing data contribute to our understanding of the potential sources and effects of selenium increases from those sources. The overview also includes a summary of the relationship of the monitoring element to conceptual and numeric models of selenium uptake and processes leading to potential impairment.

B. **Goals of the Monitoring Program**
The goals of the monitoring program are presented in a short narrative of how the monitoring results will be used in meeting the objectives of the overall monitoring program.

C. **Sampling Design**
The sampling design is a technical description that provides a summary of the recommended monitoring program at a level of detail that could serve as a basis for the development of a sampling plan. This includes:
   - individual parameter measurements
   - location(s)
   - sample frequency
   - sample replication, and
   - laboratory analyses

D. **Cost of the Monitoring Program**
An estimated annual budget for performing the work at the targeted level is presented. This includes costs for field work, selenium analyses, data management and a brief sampling report. While the costs for each element are reported separately, it is assumed that the selenium analyses will benefit from the sampling analysis for the other parameters at selected locations.

E. **Monitoring Program Performance**
A statistical perspective is critical to the development of a successful monitoring program. The monitoring program must be effective (produce information of the necessary quality), efficient (produce it at a reasonable cost) and feasible (Gitzen et al., 2012). Keys to addressing these requirements include a detailed description of the data

analyses, the statistical methods that will be used, and an analysis of the expected performance (i.e., determination of the probability of detecting meaningful change with alternative monitoring designs). For each indicator, the existing data are used to evaluate alternative study designs and analysis methods.

F. Statistical Test Methods and Plans for Analysis of the Data
Evaluation of the monitoring program performance will require the evaluation of alternative statistical tests and levels of sampling effort. The results of these analyses will be used to specify the test methods and recommended reporting requirements to support the decision-making process.

The monitoring programs developed by the subgroups are described below and summarized by these six elements in Section 7 of this report.

3 Organization of the Report

Sections 4 – 6 summarize the development and the recommendations for each of the three individual indicators. Section 7 provides a summary and a set of recommendations for moving forward with the development of an enhanced, integrated selenium monitoring plan in NSFB and the Delta to track future changes in selenium loading to the system.

4 Evaluation of Monitoring Strategies for Water-column Selenium Concentrations

Analyses conducted in 2015 in support of the TMDL and implementation plan for selenium in NSFB provided a synthesis of different monitoring efforts to characterize selenium concentrations in water in the Estuary. For water quality evaluation in the TMDL, the sources of data included studies of selenium speciation across the estuarine salinity gradient, performed in 1999-2000 and again in 2010 and 2012 (Cutter and Cutter, 2004, Doblin et al., 2006, Tetra Tech, 2012), as well as samples collected by the Regional Monitoring Program for Water Quality in San Francisco Bay (RMP). There is also regular and continuing sampling performed by the U.S. Geological Survey (USGS) at the riverine inputs to the Delta on the Sacramento and San Joaquin Rivers (with acceptable detection limits after 2007), and these data were used in recently completed TMDL-related loading analyses (Tetra Tech, 2017a).

There is currently no systematic selenium sampling program for the water column in either the Delta or the Bay; this data gap is addressed through the present analysis, which is supported by the data referenced above and by a limited period of sampling performed by the USGS (Stewart, unpublished data).

As part of this effort, and to serve as a basis for evaluating future monitoring results, a memorandum was prepared to evaluate the most recent changes in selenium in the NSFB and Delta (post-2012) through analysis of available data and modeling (Tetra Tech, 2017a). The following information is relevant to the design of future monitoring efforts:

1. **Analysis of Water-column Selenium Data in the Delta and Riverine Inflows, Collected Over the Past Decade.**

   For the Sacramento River at Freeport, dissolved selenium concentrations in recent years (2008-2017) have varied within a narrow band from 0.05 µg/L to 0.2 µg/L. Selenium concentrations seem to exhibit seasonal variability within this narrow band: higher selenium concentrations are associated with higher chloride concentrations, indicative of water evaporation/loss processes in the watershed as opposed to any change in the magnitude of upstream sources.

   Selenium concentrations in the San Joaquin River at Vernalis have shown decreases after 2011. The decreases may be due to implementation of the Grassland Bypass Project, which has lowered selenium discharges to the river in the watershed upstream from Vernalis. Reductions in Vernalis loads have an important effect on loads to the NSFB, and are also relevant to long-term outcomes from the implementation of the WaterFix project (California Waterfix, 2018).

2. **Model Evaluation of Changes in the Bay because of Changing Inflows.**
   During recent dry periods, relatively high concentrations of selenium in the Bay have been reported. One hypothesis for elevated selenium found in the Bay during dry years (such as 2014 and 2015) is the longer residence times during lower flows, allowing for greater accumulation of point and non-point loads delivered to the Bay. This hypothesis was tested using the updated ECoS 3 model, calibrated to the estuarine selenium transect data. To test the impacts of the decreased inflow to the Bay, a scenario of decreasing inflow to the Bay was run by decreasing Sacramento River flow inputs by 80%, with the other point source loads and flows remaining the same. The results at Carquinez Strait suggest that with a substantial decrease in Sacramento River flow, like what occurred during the recent drought, selenium concentrations at Bay locations could increase by up to 0.05 µg/L.

3. **Model Evaluation of the Relative Mix of Riverine Sources of Selenium that Reach the Bay.**
   The Delta Simulation Model (DSM2) is widely used to estimate flow and water quality conditions in the Delta. DSM2 modeling of the Delta suggests that San Joaquin River volumetric contribution to the Delta locations is higher during wet years. ECoS modeling of the Bay suggests higher selenium concentrations when overall freshwater flows are low. Different mechanisms therefore apply in different seasons and water year types: high concentrations during high flows may be associated with a greater San Joaquin contribution, and high concentrations during low flows may be a consequence of longer residence times of selenium in the Bay water.

4. **Updates to the Riverine Selenium Loads Delivered to the Bay.**

   In the near term, factors that may cause selenium concentrations in the NSFB and Delta to change include concentration changes in the inflow from the San Joaquin River basin, changes in refinery inputs, changes in stormwater and tributary loads from the Bay margin, and changes in overall Central Valley hydrological conditions, such as the extreme wet and dry periods that occurred between 2012 and 2017. Other factors, such as nutrient concentrations and algal levels, may also play a role, especially for the concentrations of selenium on or in particulates.

Over the longer term, selenium changes may occur due to modification of Delta flows and the mix of riverine sources because of the implementation of the WaterFix project. In a separate analysis (Tetra Tech, 2017b), The DSM2 model was used to predict volumetric contributions from six source boundaries to Mallard Island, under the altered hydrological conditions in the preferred CEQA alternative for the WaterFix project (Alternative 4). The results show increased selenium concentrations from existing conditions in the range of 8 – 20%. This range is used in the analyses presented below as a potential level of change to detect in a long-term modeling program.

A key conclusion from the most recent evaluation of water-column selenium concentration changes in the NSFB and Delta is that continued monitoring in the Bay, across all water year types, will provide an increased understanding of the controlling mechanisms, and provide insight into selenium exposures under different hydrologic conditions and seasons. This information enhances the ability to interpret the changes that are measured in biota such as sturgeon and bivalves that are planned to be monitored in the Bay. For example, the water quality modeling/monitoring effort can help evaluate whether changes in biota are the result of the hydrologic variability in the system, or are caused by a change in the system, such as loading levels, operational changes, or new infrastructure.

The above analyses provide the following points of reference for the detection of future changes in water-column concentrations:

- The expected changes in Bay selenium concentrations related to WaterFix are a consequence of change in the relative proportion of San Joaquin River versus Sacramento River water that reaches the Bay, and not just its concentration. Thus, the monitoring needs to include one or more stations downstream of the confluence of the Sacramento and San Joaquin Rivers.
- There is large year-to-year hydrologic variability in the system, as documented by the century-long streamflow record. This is associated with a large background variability in selenium loads and Bay concentrations.
- The year-to-year variability in Bay selenium concentrations, driven by hydrology, is expected in some cases to be larger than the modeled change due to WaterFix implementation in the Delta.
- Any system-wide change in selenium concentrations must be detected after correcting for the effects of hydrologic variability.

4.1 Monitoring Program Design Analyses

Three sets of analyses to support the development of a monitoring strategy capable of detecting specified levels of change in water-column selenium concentrations in NSFB are presented below. The first set was presented at the July 2016 Selenium Workgroup technical workshop. The results of those analyses provided a starting point and a basis of comparison for the identification and evaluation of additional methods that could lead to the development of more effective and efficient monitoring efforts. The primary design effort was evaluation of the applicability and expected performance of a Change Point Model (CPM) for detecting a change in water-column selenium concentrations. In the third set of analyses, historical data were used to define the "normal" or expected ranges for measured selenium concentrations in water. Then deviations from those ranges were evaluated as indicators of change.

4.1.1 Regression Methods

The July 2016 Selenium Workgroup technical workshop featured a discussion of the importance of considering statistical power in the design of temporal trend monitoring, where power is defined as the probability that the dataset of interest is sufficiently sensitive to detect a change or trend of a specific magnitude. A point was made that it is essential that power be evaluated so that the reliability of datasets, which provide a basis for assessment and regulatory action, is known. As the power increases, the probability of incorrectly concluding that no significant change has occurred will decrease. Thus, in an effective monitoring program the power should be as high as possible.

A preliminary set of statistical power analyses was presented to convey the power concept and the tradeoffs between statistical power, level of sampling effort, and sampling duration. These analyses were based on a basic understanding of the data characteristics of the three indicators of interest and of statistical distribution of environmental parameter values in general.

For trend monitoring sampling, statistical power is determined by the monitoring design parameters; these are defined by the level of sampling effort and the population characteristics of the indicator variable. The preliminary set of power analyses, described below, considered monitoring designs for water-column selenium concentrations, but the general conclusions also apply to the other monitoring parameters. The following design parameters were included:

- Level of change in water-column selenium concentrations. Two change levels were selected based on modeling results: 10% change over 10 years, an annual percent change (APC) of 1%; and 20% change over 10 years, APC = 2%.
- Variability of selenium concentrations in the sampling medium. Three levels of variability in the monitoring-parameter population were considered, corresponding to low to medium sample variability. The variability was specified as a coefficient of variation (ratio of the population standard deviation to the population mean). The selected values were 0.2, 0.3 and 0.4.

- Sampling effort. Twelve different levels of sampling effort were simulated using a combination of the number of samples per year (10, 20, and 40) and the number of years sampled (5, 10, 15, and 20).

For each set of design parameters, 10,000 sampling events were simulated. For the individual simulations, a *t* test was conducted to test the significance of the slope coefficient from the linear regression trend line. The proportion of significance test results in the 10,000 simulations provided the estimate of the statistical power (the probability of detecting the simulated trend). The higher the power, the greater the likelihood of detecting a trend.

4.1.1.1 Results

The calculated power for each simulated case is presented in Table 4-1. The levels of power are color-coded as ≥ 0.8 (green), 0.5 – 0.8 (yellow), < 0.5 (uncolored); results are summarized as follows:

- Probability of detecting the smaller simulated trend, 10% over 10 years
  - At the lowest level of variability (0.2 CV), collection of 40 samples per year is required to be reasonably confident of detecting this level of change within 10 years (Simulation 10, power = 0.89).
  - At the highest level of variability (0.4 CV), collection of 40 samples per year is required to be reasonably confident of detecting this level of change within 15 years (Simulation 35, power = 0.84). At this same level of sample variability, these analyses show that the collection of 10 samples per year for up to 20 years would not provide confidence of being able to detect the 1 % per year level of changes (Simulations 25-28).
- Probability of detecting a larger trend, 20% over 10 years
  - At the lowest level of variability (0.2 CV), the collection of only 10 samples per year is required to be reasonably confident (power = 0.88) of detecting this level of change within 10 years (Simulation 38).
  - At the highest level of variability tested (0.4 CV), the collection of 20 samples per years is required to be highly confident (power = 0.97) of detecting this level of change within 15 years (Simulation 67).

4.1.1.2 Summary

These preliminary analyses illustrate the general effects of the design parameters on the expected performance of monitoring programs. The fundamental conclusion from the values used in these simulations is that a long-term commitment to monitoring efforts is required to establish a high level of confidence to guide the assessment process. However, there are some key features associated with these preliminary analyses that suggest the need for more detailed and more realistic analyses of monitoring strategies. For example, while an expected range of sample variability was captured in the simulations, the level of variability was held constant over time. But we know that the level of sample variability for water-column and bivalve-tissue selenium concentrations changes year-to-year and month-to-month due to changes in Delta flows and the complex effects on hydrology. In the analyses presented below in this section and the next two sections, more detailed analyses of the expected performance of the proposed monitoring efforts are presented. These analyses incorporate information on our knowledge of the sampling environment, and more sophisticated and appropriate statistical methods are applied in the analyses.

Table 4-1. Preliminary Power Analysis

TABLE 4-1. Preliminary Power Analysis Results

Simulation	CV	Samples /year	Number of years	Annual Percent Change	Power	Simulation	CV	Samples /year	Number of years	Annual Percent Change	Power
1	0.2	10	5	1	0.12	37	0.2	10	5	2	0.26
2			10		0.42	38			10		0.88
3			15		0.84	39			15		0.99
4			20		0.99	40			20		0.99
5	0.2	20	5	1	0.17	41	0.2	20	5	2	0.41
6			10		0.65	42			10		0.99
7			15		0.98	43			15		0.99
8			20		0.99	44			20		0.99
9	0.2	40	5	1	0.25	45	0.2	40	5	2	0.63
10			10		0.89	46			10		0.99
11			15		0.99	47			15		0.99
12			20		0.99	48			20		0.99
13	0.3	10	5	1	0.09	49	0.3	10	5	2	0.16
14			10		0.25	50			10		0.61
15			15		0.53	51			15		0.97
16			20		0.84	52			20		0.99
17	0.3	20	5	1	0.12	53	0.3	20	5	2	0.23
18			10		0.38	54			10		0.85
19			15		0.79	55			15		0.99
20			20		0.99	56			20		0.99
21	0.3	40	5	1	0.16	57	0.3	40	5	2	0.38
22			10		0.6	58			10		0.99
23			15		0.97	59			15		0.99
24			20		0.99	60			20		0.99
25	0.4	10	5	1	0.08	61	0.4	10	5	2	0.13
26			10		0.18	62			10		0.41
27			15		0.37	63			15		0.85
28			20		0.65	64			20		0.99
29	0.4	20	5	1	0.1	65	0.4	20	5	2	0.17
30			10		0.26	66			10		0.64
31			15		0.59	67			15		0.97
32			20		0.9	68			20		0.99
33	0.4	40	5	1	0.13	69	0.4	40	5	2	0.26
34			10		0.41	70			10		0.89
35			15		0.84	71			15		0.99
36			20		0.99	72			20		0.99

4.1.2 Change Point Detection Method

4.1.2.1 Method Description

An evaluation of the effectiveness of a Change Point Model (CPM) in detecting whether a change in water-column selenium concentrations has occurred over a specified time interval was conducted. The analyses were conducted using the **R** package **cpm** described by Ross (2015). The cpm package provides an effective and computationally efficient method for sequential detection of multiple change points in sequences of random variables $x_1, \dots, x_n$ and is especially well suited to the data time series of interest.

The method is based on the sequential calculation at each time, t, of the test statistic

$$D_t = \max_{k=2,\dots,t-1} D_{k,t},$$

where $D_{k,t}$ is a generic two-sample hypothesis test-statistic that a change in distribution occurred at time k. In this analysis, we test the power of two different $D_{k,t}$ alternatives: a Student t statistic for detection of a change in the mean of a sequence of normal variables and a Mann-Whitney statistic for general location shift in an unspecified distribution. If $D_t > h_t$ for an appropriate threshold value h_t, then a change is determined to have occurred in the sequence $x_1, \dots, x_t$, and the change is determined to have happened at the value of k that maximizes $D_{k,t}$. If D_t does not exceed the threshold, the process is continued upon receiving the next data point, x_{t+1}. Generally, h_t is chosen such that the probability of a Type I error is constant over time and is parameterized in terms of the average run length ARL_0 (the average number of observations received before a detection is flagged under the null hypothesis of no change).

4.1.2.2 Steps in Conducting the Analysis

Because water-column selenium data for the Bay are limited, the following approach was used to create a synthetic time series that was used in testing for an effective monitoring program in conjunction with a change detection methodology:

- Use data with mean and standard deviation from the USGS Bay data (Stewart, unpublished data, 2015-2017)
- Impose hydrologic change in the following manner on selenium concentrations: wet years (-40% of mean); above normal years (-20% of mean); below normal years (mean); dry years (+20% of mean); and critically dry years (+40% of mean)
- Add an increase of mean concentrations of either 8% or 20%, implemented over 10 years, as a representation of a WaterFix-induced change. The range of increase is based on modeling summarized above. The 10-year duration of change is a reasonable representation of the time frame that a project of this magnitude would need to be fully operational.

We had access to about three years of water quality data from three different sites (Stewart, unpublished data), with approximate mean of 0.12 µg/L, standard deviation of 0.024 µg/L, and autocorrelation of 0.4. An underlying assumption is that recommended sampling stations identified in Section 4.1.4.1 (Sampling Design) exhibit the same variability as these three USGS sampling sites.

To estimate the power of this method on data like the observed data, we generated 37 years of monthly autocorrelated random normal data and then introduced year type variation into the simulated data by applying the above adjustments corresponding to the sequence of historical water year types from 1980-2017. The mean, variance, and autocorrelation of this generation process were adjusted such that the last 36 months of random data (every-other month, as the available data were bimonthly) had properties like the USGS dataset (see Figure 4-1).

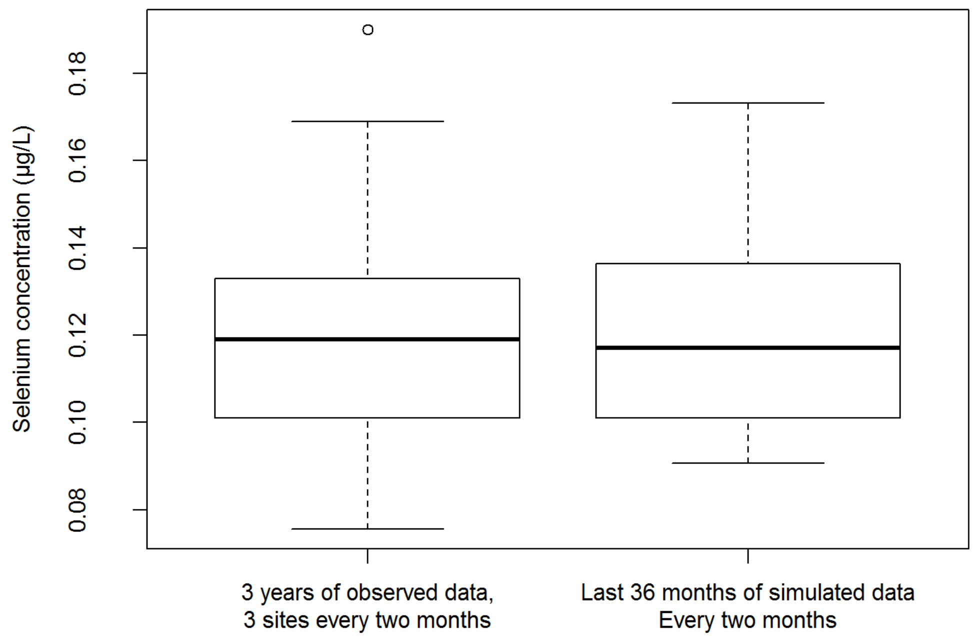

Figure 4-1. Comparison of observed and simulated data using boxplots. Observed data are from Stewart (Unpublished).

Two distinct levels of change were imposed on these synthetic datasets based on modeling results: (1) 8% increase over 10 years, i.e., a 0.10 proportional change or 0.8% of the overall mean concentration per year, and (2) 20% increase over 10 years, i.e., a 0.10 proportional change or 2% of the overall mean concentration per year. In both cases, the trend began at year 5 and ended at year 15 in the simulated datasets, with the underlying level being held constant before and after that, as seen in Figure 4-2.

The test was evaluated with three different levels of the ARL_0 parameter described above: 500, 1000, 10,000. We also tested keeping only every 2nd, 4th, or 6th sample before running the CPM test, corresponding to 6, 3, and 2 samples per year, respectively.

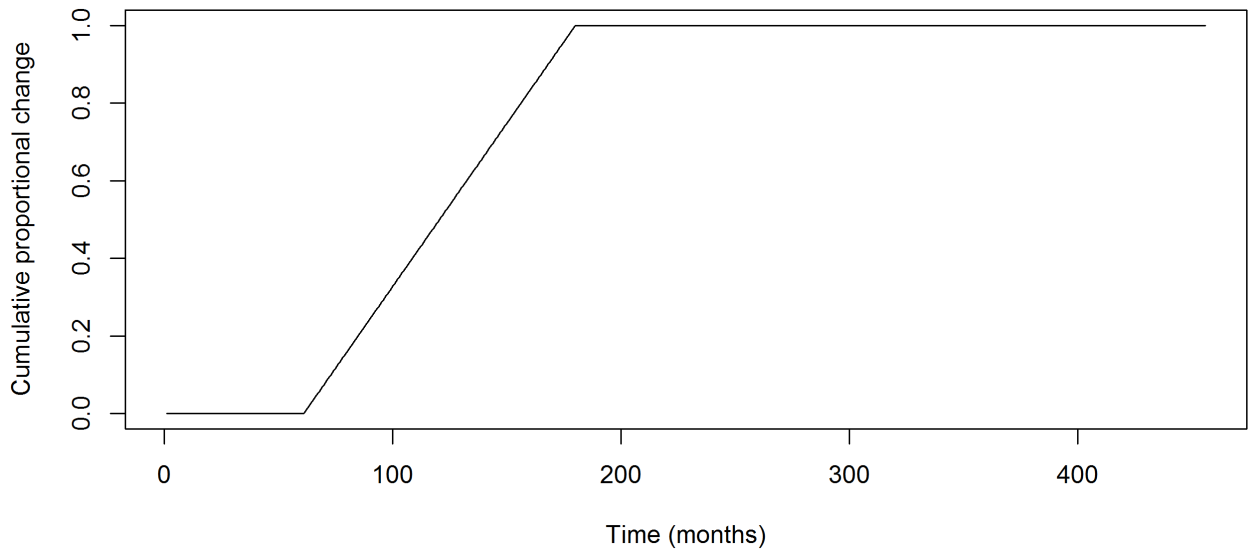

Figure 4-2. Imposed change in concentrations over a 10-year period, from year 5 through 15 of the synthetic record of 1980-2017.

4.1.2.3 Results

The sequence of applying the method to one replicate of the simulated data is shown in Figure 4-3. The CPM method receives the data generated by the method described above. The synthetic data series are generated multiple times to perform the test.

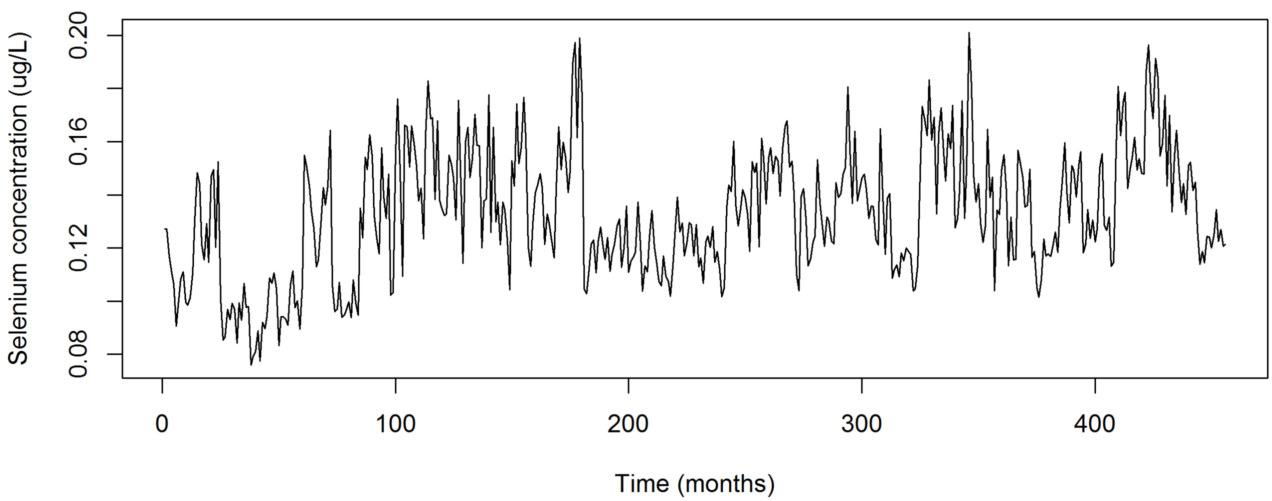

Figure 4-3. Monthly raw data generated over 37-year period (1980-2017), single realization of synthetic time series

These data are then normalized (corrected) by water year type (Figure 4-4).

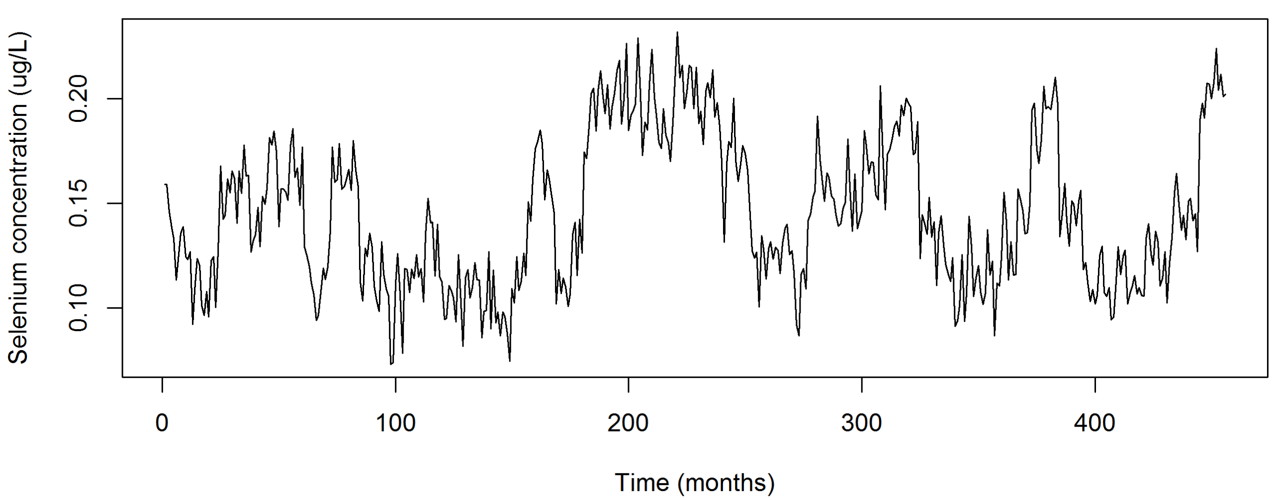

Figure 4-4. Raw data in Figure 4-3 normalized by water year type (actual occurrence of water year types over 1980-2017)

Autocorrelation is removed by estimating an autoregressive model, AR(1) with linear trend (red) (Figure 4-5).

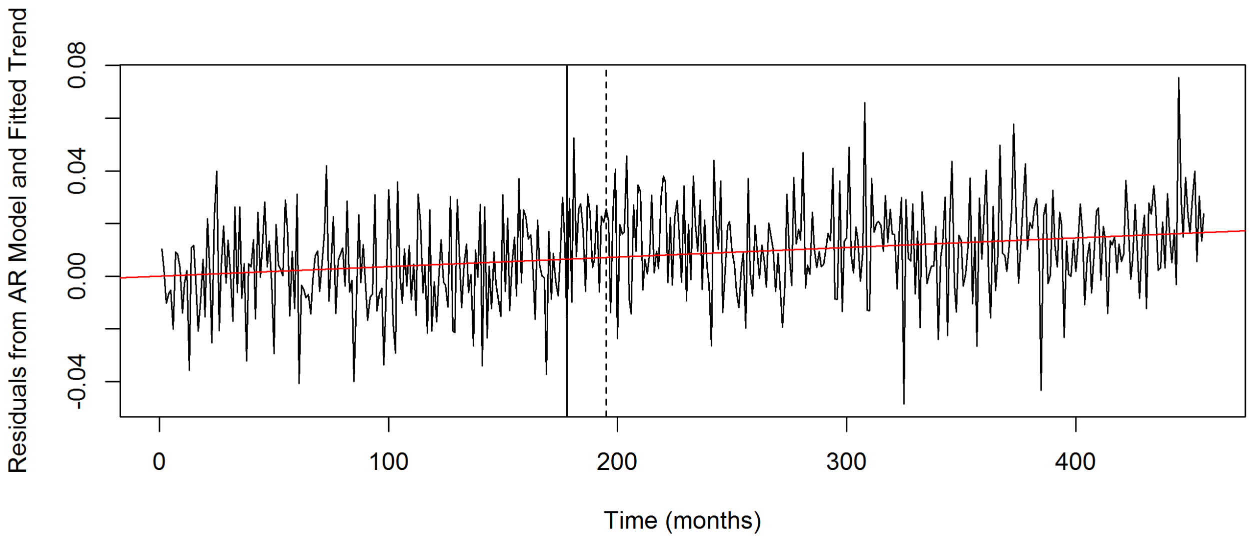

Figure 4-5. Residuals in time series following application of AR model with trend to data in Figure 4-4

The CPM method evaluates the residuals from this model with the linear trend reintroduced. In Figure 4-5, the dashed vertical line is the time t at which $D_t > h_t$, i.e. the detection time, and the solid vertical line is at the time k for which the value of $D_{k,t}$ was maximized, i.e. the estimated change point. Results are shown in Table 4-2 with simulations performed multiple times.

For the smaller imposed change (8% over 10 years), no combination of ARL_0, sampling frequency or test statistic gave strong values of estimated power. When the change was detected (50% of the time), detection delays from the start of the trend at month 60 ranged from 113-144 months (yellow highlights in Table 4-2).

For the larger change (20% over 10 years), 6 or 12 samples per year were sufficient to give good power (>80% of detection) for all combinations ARL_0 and test type (Student's t or Mann-Whitney test statistics), with detection times ranging from 106 to 179 months (green highlights in Table 4-2). A lower level of power (>50%) was achievable with 2 or 3 samples per year at ARL_0 of 500 or 3 samples per year with $ARL_0 = 1000$, with longer detection times of 164-182 months (yellow highlights).

Table 4-2. CPM Results Summary

ARL_0	Change Size	$D_{k,t}$ test type	Months between samples	Power estimate*	Average detection delay (months)
500	0.08	Mann-Whitney	1	0.63	119
500	0.08	Mann-Whitney	2	0.44	142
500	0.08	Mann-Whitney	4	0.24	157
500	0.08	Mann-Whitney	6	0.14	191
500	0.08	Student	1	0.6	113
500	0.08	Student	2	0.43	137
500	0.08	Student	4	0.21	154
500	0.08	Student	6	0.12	166
1000	0.08	Mann-Whitney	1	0.52	144
1000	0.08	Mann-Whitney	2	0.33	169
1000	0.08	Mann-Whitney	4	0.16	183
1000	0.08	Mann-Whitney	6	0.1	187
1000	0.08	Student	1	0.49	142
1000	0.08	Student	2	0.31	160
1000	0.08	Student	4	0.13	172
1000	0.08	Student	6	0.073	181
10000	0.08	Mann-Whitney	1	0.33	212
10000	0.08	Mann-Whitney	2	0.13	217
10000	0.08	Mann-Whitney	4	0.032	192
10000	0.08	Mann-Whitney	6	0.011	199
10000	0.08	Student	1	0.28	216
10000	0.08	Student	2	0.091	219
10000	0.08	Student	4	0.029	212
10000	0.08	Student	6	0.014	213
500	0.2	Mann-Whitney	1	0.99	107
500	0.2	Mann-Whitney	2	0.96	132
500	0.2	Mann-Whitney	4	0.84	161
500	0.2	Mann-Whitney	6	0.64	182
500	0.2	Student	1	0.98	106
500	0.2	Student	2	0.95	133
500	0.2	Student	4	0.79	164
500	0.2	Student	6	0.6	185
1000	0.2	Mann-Whitney	1	0.98	119
1000	0.2	Mann-Whitney	2	0.95	144
1000	0.2	Mann-Whitney	4	0.77	173
1000	0.2	Mann-Whitney	6	0.5	190

ARL_0	Change Size	$D_{k,t}$ test type	Months between samples	Power estimate*	Average detection delay (months)
1000	0.2	Student	1	0.96	122
1000	0.2	Student	2	0.91	144
1000	0.2	Student	4	0.74	175
1000	0.2	Student	6	0.48	195
10000	0.2	Mann-Whitney	1	0.97	146
10000	0.2	Mann-Whitney	2	0.87	174
10000	0.2	Mann-Whitney	4	0.47	210
10000	0.2	Mann-Whitney	6	0.21	226
10000	0.2	Student	1	0.96	148
10000	0.2	Student	2	0.82	179
10000	0.2	Student	4	0.37	212
10000	0.2	Student	6	0.15	235

* Green highlight: power ≥ 0.80; yellow highlight: 0.5 < power < 0.80

4.1.2.4 Implications for Monitoring Selenium in the Water Column

The above application of change detection with synthetic data highlights the challenge of finding change in an environment with large background variability. The results show that for a reasonably small change (~8%), sampling with monthly frequencies will be insufficient for robustly detecting change. For a larger change (~20%), monthly sampling could detect change over a period of several years. Bi-monthly sampling (once every two months), reduces the power somewhat, but still provides reasonable basis to detect change over about 10-15 years.

Of course, we do not know *a priori* what the magnitude of change will be, but given a range as above, can commit to a certain level of sampling. Assuming that 6 or 12 samples per year are reasonable, the testing shows that changes smaller than about 20% would be difficult to detect. Alternatively, the sampling frequency can be thought of as a way of checking whether a change greater than 20% has occurred.

4.1.3 Normal Range Statistical Testing

The third approach evaluated for detecting changes in selenium concentrations that warrant management actions is the comparison of monitoring data with “normal ranges”. Normal ranges are defined as some fraction of a reference distribution that is considered representative of acceptable or “normal” conditions. Direct comparison of measured values in a monitoring sample to the limits of the normal range can be used to detect changes. For example, in Figure 4-6, the normal range is represented by the shaded area of the distribution that encompasses 95% of the previous observations. The occurrence of a new sample or new mean value from the monitoring program that falls outside the designated normal range raises a yellow flag that warrants additional investigation.

These range comparisons can be formalized with the use of a *t* test of the differences between two mean values (e.g., the means of reference conditions [reference, $\bar{x}_r$] and existing monitoring conditions [$\bar{x}_e$]). The expressions on the left-hand side of Equations [1] - [3] are described by the central *t* distribution:

$$\frac{\bar{x}_r - \bar{x}_e + \mu_r - \mu_e}{S_{\bar{X}}} = t \qquad \textbf{[1]}$$

where

- $S_{\bar{X}}$ is the pooled standard error (SE) for the difference in means
- μ_r and μ_e are the parametric means

Rearranging Equation [1]:

$$\frac{\bar{x}_r - \bar{x}_e}{S_{\bar{X}}} + \frac{(\mu_r - \mu_e)}{S_{\bar{X}}} = t \qquad \textbf{[2]}$$

Under the null hypothesis that the two samples come from the same population and have the same parametric mean ($\mu_r - \mu_e = 0$), the expression for t is given by:

$$\frac{\bar{x}_r - \bar{x}_e}{S_{\bar{X}}} = t \qquad \textbf{[3]}$$

The t test is conducted to evaluate for the deviation of the difference, $\bar{x}_r - \bar{x}_e$, from 0.

Figure 4-7 shows the central *t* distribution for 18 *df.* The critical values for $\alpha = 0.05$ are $\pm$ 2.1, and the area between the critical values is shaded in blue. If the quantity $\frac{\bar{x}_r - \bar{x}_e}{S_{\bar{X}}}$ is outside the shaded area defined by critical *t* values ($\pm$ 2.1), we conclude that the difference between the means is statistically significant (i.e., the difference in the means is outside the expected range for the sample size and sample variability).

In the case that $\mu_r \neq \mu_e$, the second term in Equation [2], $\frac{(\mu_r - \mu_e)}{S_{\bar{X}}}$, has a non-zero value and is designated the non-centrality parameter, δ, and the non-central t distribution, *t'*, is distributed around δ. By specifying the difference between μ_r and μ_e we can determine the value of the non-centrality parameter, define the *t'* distribution, and determine the power of a *t* test (i.e., the ability to detect differences the means of reference conditions [reference, $\bar{x}_r$] and existing monitoring conditions [$\bar{x}_e$]).

The determination of the power calculation is shown graphically in Figure 4-8. The central *t* distribution from Figure 4-7 is shown on the left-hand side of Figure 4-8. The non-central *t'* distribution for a specified level of difference between the reference and effect mean values, $\bar{x}_r - \bar{x}_e = \Delta$, is shown on the right-hand side (RHS) of Figure 4-8. The blue vertical line in Figure 4-8 is located at the upper 95th percentile of the central distribution. The gray-shaded area within the non-central t distribution (t') on the RHS of Figure 4-8 is the critical region for the non-central null hypothesis. The area under the curve in the shaded area represents the probability of obtaining a significant t-test result when $\bar{x}_r - \bar{x}_e = \Delta$. By specifying different values for Δ, we can estimate the power of the *t* test under different test conditions.

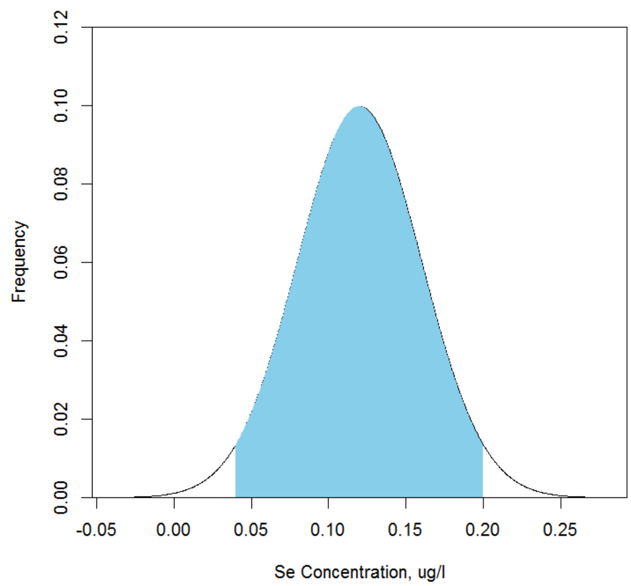

Figure 4-6. Normal range representing 95% of the expected distribution.

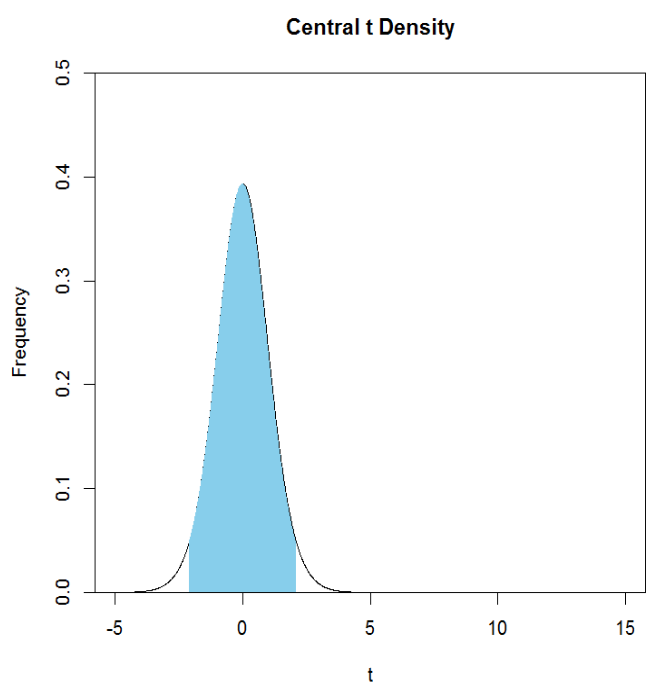

Figure 4-7. Central t distribution for 18 df.

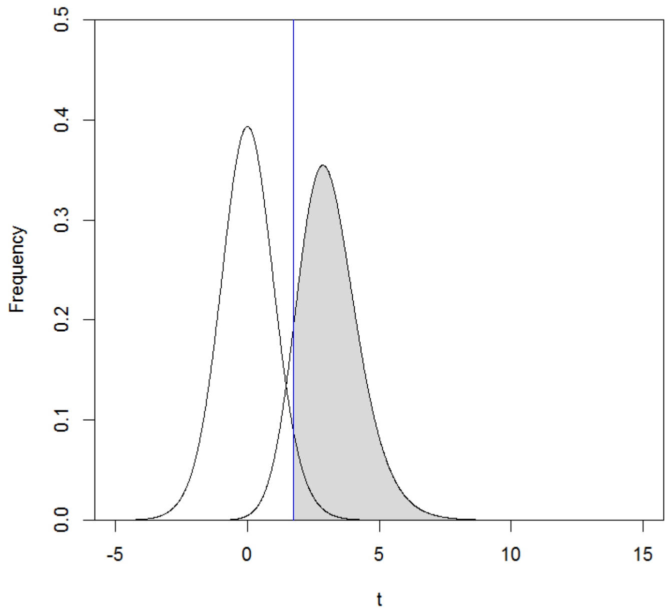

Figure 4-8. Central and non-central t distributions for 18 df.

Using the estimates of the mean and standard deviation developed in the above for the evaluation of the Change Point Detection method ($\bar{x} = 0.12$ and s = 0.024), power analyses were conducted to evaluate the power of *t* tests in detecting deviations from the distribution of values based on the historical record. The results are presented in Table 4-3 and show the effect of sample size and percent change in the year-to-year mean concentration on the probability of detection. These results suggest that there is a high level of probability of detecting changes in the mean concentration of 30 – 40 % with 20 samples per year or of detecting a 40 % change with 10 samples per year (green shaded cells).

Table 4-3. Normal Range Power Analysis Results

Samples/year	Percent Change	Power
10	20	0.31
20	20	0.49
10	30	0.58
20	30	0.84
10	40	0.88
20	40	0.99

4.1.4 Proposed Monitoring Framework for Water

Goals of the Monitoring Program. The TMDL and implementation plan for selenium in NSFB and the USEPA's proposed aquatic life and aquatic-dependent wildlife criteria for selenium in the

Bay and Delta established targets for dissolved water-column selenium. Some level of water quality monitoring will be required to assess compliance with these targets, regardless of conditions in the estuary and its contributing watershed. However, the specific goal of the sampling strategies evaluated in this document is to establish a monitoring plan that can detect increases in water-column selenium concentrations that warrant prompt management responses to signs of increasing risk of impairment.

4.1.4.1 Sampling Design

- Primary parameters: Total, dissolved and particulate selenium. Particulate selenium directly measured on filters, and not by difference between total and dissolved.
- Ancillary data: total suspended material, total organic carbon, chlorophyll *a*.
- Locations: Stations in Suisun and San Pablo Bays and the Delta (monitoring of the Delta stations is not presumed to be funded through the San Francisco Bay RMP). For stations in the Bay, co-located with clam sampling stations at Mallard Island (Station 4.1) and the Carquinez Strait (Station 8.1). New Delta stations proposed as shown in Figure 4-9. Four sampling stations on tributaries/sloughs before entering the Bay are recommended. These stations include the Sacramento River at Rio Vista (RIO), San Joaquin River at Jersey Point (JER), Three Mile Slough (TMS), and Dutch Slough (DCH). These four rivers/sloughs represent major pathways entering the Bay before their confluence at Mallard Island. Three Mile Slough connects between the Sacramento River and the San Joaquin River. Dutch Slough joins San Joaquin River upstream of San Joaquin River at Antioch. Water quality data, such as temperature, specific conductance, nutrients, and suspended sediments, are available for these stations. The Mallard Island station (Station 4.1) is most downstream of the Delta, below the confluence of the Sacramento River and the San Joaquin River and is the same station used in the USGS long-term bivalve studies. Mallard Island represents results of mixing from all sources of selenium in the Delta. Two stations in the interior Delta with extensive existing water quality data are also proposed to monitor contributions from the San Joaquin River to the Delta: Old River at Bacon (OLD) and Middle River at Holt (MRC). Fingerprint studies at these stations showed noticeable contributions from the San Joaquin River, particularly during wet months and wet years (Tetra Tech, 2017b). It is assumed that the two riverine stations at Freeport and Vernalis, will continue to be monitored through ongoing programs, as done presently, and costs for sampling the riverine locations are not included in this design.
- Sample frequency: Monthly for three years. Then assess monitoring program performance and evaluate projected effectiveness of bi-monthly (once every two months) sampling.
- Replication: Single monthly grab samples collected near the surface.

4.1.4.2 Cost of the Monitoring Program

- RMP monitoring of water will only cover Stations 4.1 and 8.1. Water and bivalves will be monitored together, with a combined average annual cost of $93,000 per year.

4.1.4.3 Monitoring Program Performance

- Greater than 80% likelihood of detecting a 20% change with detection times within a 10-year sampling period.

4.1.4.4 Statistical Test Methods and Plans for Analysis of the Data

- Monitoring program performance is based on the results of the Change Point Detection Method analyses. These analyses were based on conservative assumptions (e.g., an annual monotonic incremental increase in concentrations). Because of the conservative assumptions, a higher level of performance is expected.
- In addition to the Change Point Detection method, the normal range testing methods should also be applied in the analysis of the data. This will provide an additional method for the detection of the annual variation that is likely to occur with important changes to the system hydrology.

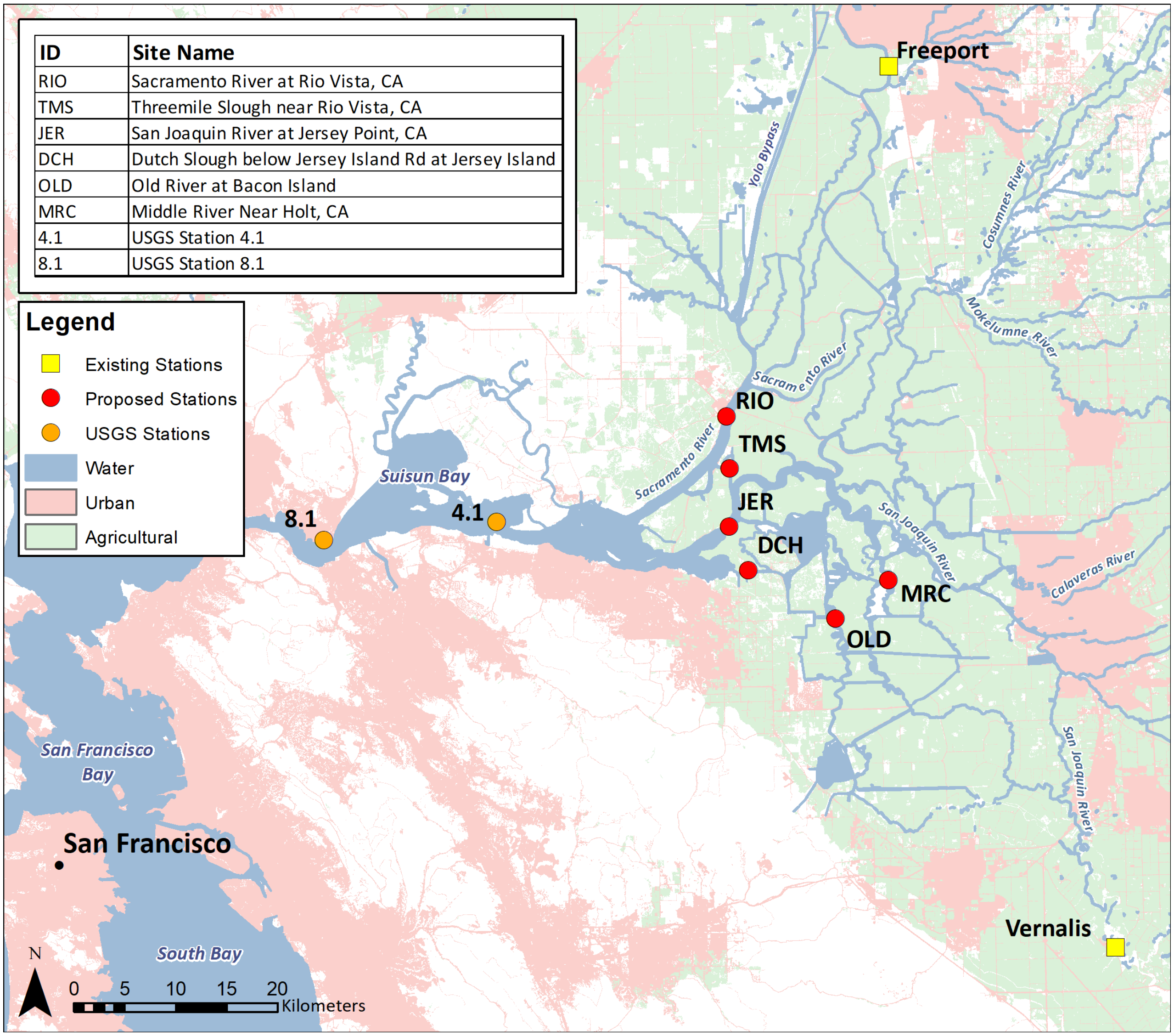

Figure 4-9. Proposed monitoring stations. Some of these stations have been monitored historically for selenium (yellow squares for riverine inputs, and orange circles associated with USGS clam sampling). A set of six new stations in the Delta is proposed (red circles). The Delta stations have not been monitored historically for selenium but are part of other regular water quality monitoring programs. Monitoring of the Delta stations is not presumed to be funded through the San Francisco Bay RMP.

5 Evaluation of Monitoring Strategies for Bivalve Selenium Concentrations

When elevated selenium concentrations were discovered in bivalve-eating predators in San Francisco Bay (Ohlendorf et al., 1986; Ohlendorf et al., 1991), USGS initiated a monthly monitoring program of selenium in benthic bivalve *Potamocorbula amurensis*. Data from this time series, extending from May 1995 through September 2017, have been published in Linville et al. (2002), updated in (Stewart et al., 2013), and are being prepared for a final summary (Stewart et al., in prep). The 22-year time series, spanning a broad range of freshwater inflows and physicochemical conditions, is unique for its duration and frequency of sampling and provides a comprehensive understanding of variation in selenium exposures at organism (i.e., clam size), site, seasonal and inter-annual levels.

USGS funding for the bivalve monitoring program was withdrawn in early 2017 and the RMP supported the continuation of sampling from April through September 2017. Sampling has not been conducted since then.

5.1 Monitoring Program Design Analyses

5.1.1 Testing against normal range

This approach is based on an adaptive monitoring process developed by Canadians as part of the Environmental Effects Monitoring programs for assessment of impacts related to mining and other industrial processes (Arciszewski and Munkittrick, 2015; Kilgour et al., 2017; Arciszewski et al., 2017). The basic premise involves identifying the expected or "normal" range of concentrations for a chemical parameter based on the distribution of a historical dataset and testing new measurements against that range. In other words, for samples that have a normal distribution we would expect any new samples collected to fall within that normal distribution range 95% of the time. Values that fall outside of this range are assumed to be random occurrences 5% of the time. Changes in selenium exposures would be identified when there are multiple concurrent exceedances of selenium concentrations (based on a mean of n sample replicates) outside the normal range, which indicate that the values are no longer occurring by pure chance.

We may also wish to consider more subtle changes, such as when selenium concentrations are increasing toward the higher end of the selenium normal distribution. In this case, we can test to see the likelihood of selenium exposures exceeding the 75^{th} percentile or the upper quartile of a data distribution.

If values exceed the 75^{th} quartile then the next step would be to ask if the values fall above the mean selenium exposure concentration ($\mu \pm 95\%$ CI) found in dry water years, which have been shown, based on historical data, to have among the highest selenium concentrations.

5.1.2 Approach:

1. Normal ranges of selenium concentrations were calculated for the 22-year time series for the following:

- **USGS Stations 4.1 and 8.1.** These stations have the longest individual time series and represent end members of the range of selenium concentrations in the Bay.
- **By collection date.** Selenium concentrations for multiple composite samples (n=2-5) of bivalves ranging in shell length from 9-16 mm are averaged for a single monthly mean. Because selenium can be affected by bivalve size, a representative selenium concentration for the population can be obtained by averaging across a range of sizes.
- **By month.** While month *per se* has no specific ecological significance, samples are collected monthly and different flow conditions are generally associated with certain months. The normal range for selenium for a given month is based on 22 years of mean monthly values (n=22).

The selenium concentrations for a given station and month do not always strictly follow a normal distribution, so estimates of the "normal" range, which can be approximated by the upper and lower bounds of 2 times the standard deviation around the mean, may not be accurate. Further, this range when estimated based on the mean (±2 standard deviations) are less precise with small sample numbers. For this reason, we estimate the normal range of values for each station and month based on the lower (2.5^{th}) and upper (97.5^{th}) quantiles, which are determined by the rank of all samples. To compute the p^{th} quantile of n non-missing values for a station and month, values are arranged in ascending order $y_1, y_2, ..., y_n$. The rank number for the p^{th} quantile is calculated as $(p / 100) * (n + 1)$. For example, for a month that has 15 samples, the Se concentration of the y_{12} value is the 75th quantile. The 90th quantile is interpolated by computing a weighted average of the 14th and 15th ranked values as $p90 = 0.6y14 + 0.4y15$ (SAS Institute Inc., 2018).

2. Selenium concentrations calculated for water years representing low flow (dry) and high flow (wet) were determined by:
   - Dividing the 22-year dataset into wet and dry years based on the designation of water year type for the Sacramento River by the California Department of Water Resources. The Sacramento water year designations were representative of the San Joaquin, except for 2003 (dry) and 2010 (wet):
     - Wet – 1995-2000, 2003, 2005-2006, 2011, 2017 (n=11)
     - Dry – 2001-2002, 2007-2010, 2012-2016 (n=11)
   - Calculating the mean ± 95% CI for each water year type, station and month.

5.1.3 Results

Results for monthly normal ranges and wet and dry mean concentrations for Stations 4.1 and 8.1 are shown in Figure 5-1. The normal ranges of selenium concentrations in *P. amurensis* are higher and larger (more variable) at Station 8.1 (range 7.55 – 16.08 µg/g) than at Station 4.1 (6.06 – 10.75 µg/g). The normal range shown for each station encompasses wet and dry water year extremes. The largest normal ranges are during the spring months (February – March) at Station 8.1 and relatively consistent across all months at Station 4.1 with slightly higher ranges in November and May. The larger ranges in the spring at station 8.1 in February and March may be due in part to the effect of low flow conditions during critically dry years resulting in elevated Se concentrations at a time when selenium concentrations are typically diluted by high flows in wet years.

5.2 Power to Detect Change

The power to detect real change (i.e., not due to random chance) in selenium concentrations is a function of the variability in selenium concentrations (standard deviation), the magnitude of change (difference in new monitoring mean from historical mean selenium concentrations), and the number of samples collected (n). The variation (standard deviation) in selenium concentrations was determined from the normal distribution for each station and each month.

The number of samples collected can be varied based on different sample designs ranging from minimal to optimal. The following three options are illustrated in Table 5-1: 1) monthly samples for 3 months, such as before fall muscle plug monitoring program (July – September) or pre-spawning period (December – February), depending on monitoring goals and priorities; 2) monthly samples for 6 months, such as the 3-month period preceding fall muscle plug monitoring and 3 months during the pre-spawning period, or alternate months throughout the year; and 3) the number of months required to be able to detect increases to at least the 97.5th quantile with 0.80 power. Power calculations for each site, month and sample number are shown in Table 5-1.

Power estimates were relatively consistent between the two stations indicating that increases in station and monthly means that are as large as the value at the upper normal range (97.5th quantile) can be detected with a power of 0.32 and 0.41 with 3 samples and with a power of 0.83 and 0.90 with 6 samples per year at Stations 4.1 and 8.1, respectively. Detecting a statistical shift above the 75th quartile with 0.8 power would require between 11 and 68 (excluding February) samples at Station 4.1 and between 11 and 24 samples at Station 8.1. Detecting very small differences from historical sample means, such as for Station 4.1 in February (n=845), requires more samples than when trying to detect a larger difference that is greater than the variation around the historical mean.

5.3 Relationships Between Bivalves and Other Sample Matrices

In the spring of 2014 the USGS developed a method for analyzing water samples for dissolved and particulate selenium concentrations (Kleckner et al., 2017). From April 2014 through September 2017, water samples were collected concurrently with bivalves. Results for selenium concentrations in water (filtered and particulates) and bivalves as well as white sturgeon muscle plugs, liver and ovary collected over the same period are shown in Figure 5-2. Variation in selenium concentrations among sample sites (represented by error bars around the mean) was lowest for filtered water followed by particulates and then bivalves (note the differing scales between water and biota). Variation among individual white sturgeon samples collected on the same sampling date was of similar magnitude for muscle plugs, liver and ovary. Sample variability, across sites or individual white sturgeon, increased from filtered water < particulates < bivalves < white sturgeon. Filtered and particulate water selenium concentrations showed slightly different seasonal patterns from each other and from the bivalves; however, most matrices (filtered water, bivalves and fish tissues) showed a similar gradual decline from the beginning of the time series to the end, while particulate selenium concentrations showed a slight increase toward the end of the time series from 2014 through 2017.

This response of decreasing selenium concentrations was not unexpected for bivalves, which historically have shown elevated selenium concentrations during periods of low flow (2014-2016) and lower concentrations during high flow (2017). The fact that this general pattern was also found in a relatively small number of filtered water and white sturgeon sampling events suggests that there is good correspondence among these environmental indicators in signaling changes in selenium concentrations in the Bay through time.

5.4 **Proposed Monitoring Framework for Bivalves**

5.4.1 Goals of the Monitoring Program

Monitoring of bivalve selenium has the following goals:

- Tracking attainment/impairment
- Providing early warning of shifts in selenium exposures
- Serving as an intermediate link between water and fish tissue criteria

Bivalves are of regulatory interest (TMDL and USEPA proposed revised criteria) because they are an important part of the diet of white sturgeon, have been shown to result in elevated exposures in bivalve predators (Stewart et al., 2004) and form an intermediate link between aqueous selenium and predator exposures. A long-term monitoring history (22-years) has been used to establish the "normal range" or expected selenium concentrations against which to evaluate changes in selenium in the future. Further, bivalves respond to their environmental relatively quickly (~30 days) and can therefore serve as early warning indicators of change.

Recent joint monitoring efforts of water, bivalves and white sturgeon indicate that there is good correspondence among these matrices (Figure 5-2). Additional analyses of stable isotopes of carbon, nitrogen and sulfur in bivalve tissue are necessary to establish a baseline from which to verify that the sturgeon continue to feed on invertebrate diets (δ^{15} N), and the general foraging ranges for the sturgeon (δ^{13} C and δ^{34} S) (see Stewart et al. 2004 and 2013 for methods and approach).

5.4.2 Sampling Design

- Parameters: Total selenium in bivalve soft tissues (species *Potamocorbula amurensis*); Carbon (δ^{13} C), nitrogen (δ^{15} N), and sulfur (δ^{34} S) isotopes
- Location: USGS Stations 4.1 and 8.1
- Sampling frequency: 6 months (alternating months, or 3 months pre-spawning and 3 months before fall sturgeon monitoring)
- Replication: 5 composites representing bivalves ranging in size from 9-16 mm.
- Analytical approach: ID-HG-ICP-MS

5.4.3 Cost of the Monitoring Program

- Water and bivalves will be monitored together, with a combined average annual cost of $93,000 per year.

5.4.4 Monitoring Program Performance

- Greater than 80% likelihood of detecting increases of at least the upper limit of the normal range (97.5th quantile) based on an average of n= 5 (Station 4.1) and n=6 (Station 8.1) monthly samples within a given year.
- Greater than 80% likelihood of detecting exceedances of the 75th quartile of the normal range with between 11 to 24 months at Station 8.1 and between 11 and 68 samples (excluding February) at Station 4.1 (based on variability for a given month).

5.4.5 Statistical Test Methods and Plans for Analysis of the Data

- Monitoring program performance is based on the normal range established using the USGS 22-year time series for *P. amurensis* at Stations 4.1 and 8.1 for individual months. Individual monthly samples can be tested against normal range bounded by the 2.5 - 97.5th quantiles, the 75th quartiles and against critical effect ranges for dry and wet water years.

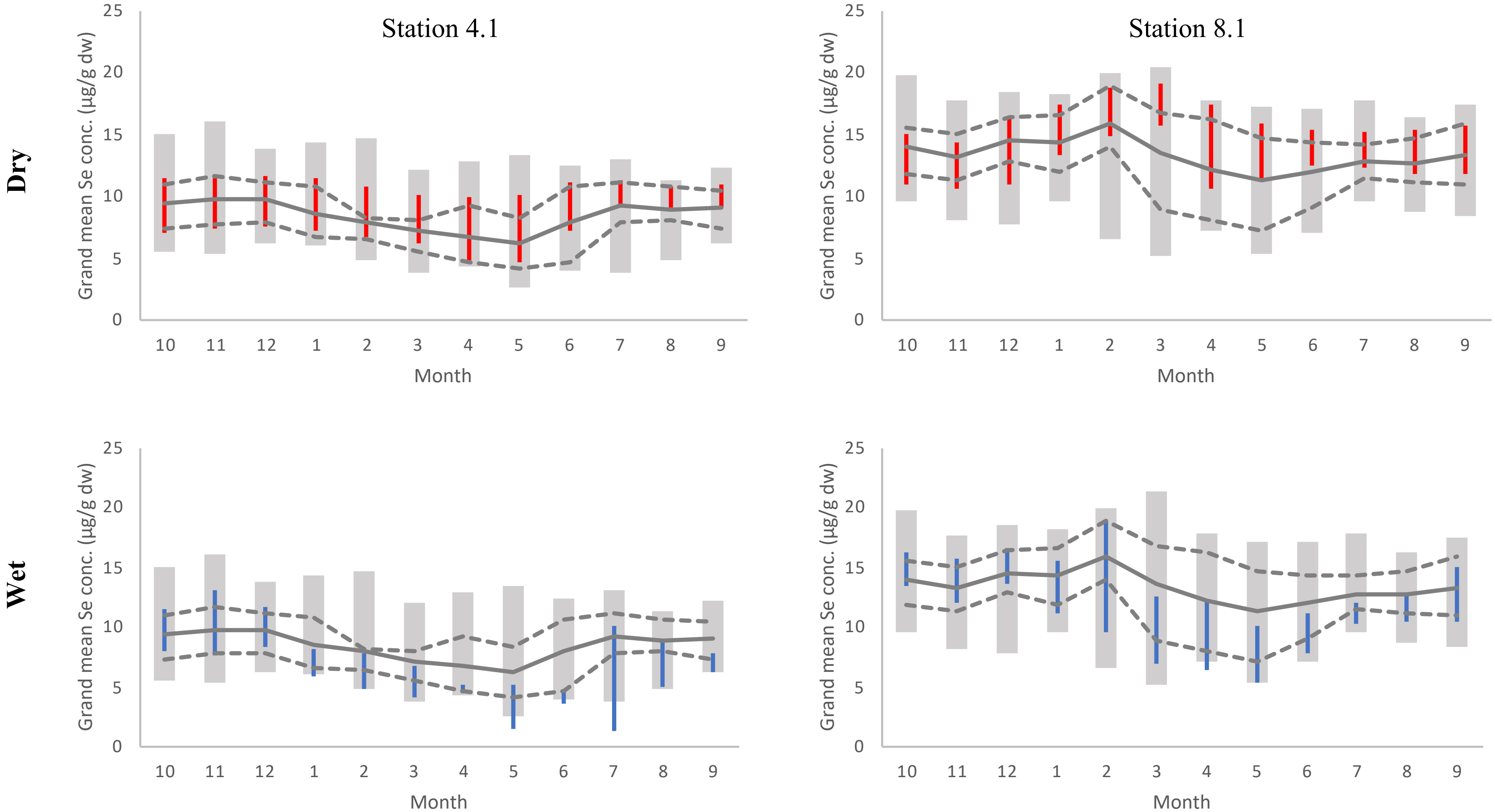

Figure 5-1. Normal ranges of monthly selenium (Se) concentrations (µg/g dry weight) in Potamocorbula amurensis based on 22 years of monitoring at two stations in northern San Francisco Bay. Grey bars – normal range defined by the upper 97.5th and lower 2.5th quantiles. Solid grey line – grand monthly mean. Dashed grey lines – interquartile range. Red vertical lines – Dry water year mean Se concentration (±95% CI). Blue vertical lines – Wet water year mean Se concentration (±95% CI).

Table 5-1. Sample and power estimates to detect increases as large as the 75th (Q75), or 97.5th (Q97.5) quantiles based on mean, standard deviation and range of Se concentrations (µg/g dw) for a given month and station across 22 years of monitoring. The number of samples needed to detect an increase in station and monthly mean Se concentration to the Q75th and Q97.5th with a power of 0.8 and the power of detecting increase to the Q75th and Q97.5th with sample numbers of 3 and 6. For example, 13 samples are needed to detect an increase in Se concentration at station 4.1 and April to the Q75th (9.18 µg/g) with 0.80 power.

Site name	Month	n	Mean Se (µg/g)	Std Dev	Q2.5th (Se µg/g)	Q25th (Se µg/g)	Q75th (Se µg/g)	Q97.5th (Se µg/)	Q75 0.8 power n	Q97.5 0.8 power n	Q75 n=3 power	Q97.5 n=3 power	Q75 n=6 power	Q97.5 n=6 power
4.1	10	16	9.47	2.47	5.55	7.38	10.97	14.96	24	4	0.10	0.54	0.23	0.99
4.1	11	16	9.80	2.74	5.33	7.74	11.65	16.00	20	4	0.11	0.55	0.27	0.99
4.1	12	16	9.73	2.26	6.20	7.81	11.13	13.83	23	5	0.10	0.41	0.24	0.94
4.1	1	14	8.56	2.50	6.05	6.64	10.75	14.25	13	4	0.15	0.56	0.41	0.99
4.1	2	15	7.92	2.59	4.79	6.45	8.17	14.72	845	4	0.05	0.65	0.05	1.00
4.1	3	14	7.15	2.49	3.83	5.47	8.00	12.11	68	5	0.66	0.47	0.11	0.97
4.1	4	11	6.71	2.80	4.35	4.70	9.18	12.87	13	4	0.15	0.53	0.42	0.99
4.1	5	11	6.21	3.29	2.60	4.06	8.27	13.35	23	4	0.10	0.52	0.24	0.99
4.1	6	12	7.91	3.17	3.95	4.59	10.70	12.45	13	7	0.15	0.30	0.42	0.80
4.1	7	14	9.22	2.46	3.77	7.81	11.09	13.00	16	6	0.13	0.33	0.33	0.85
4.1	8	15	8.92	1.86	4.77	8.01	10.67	11.33	11	7	0.17	0.26	0.46	0.72
4.1	9	16	9.11	1.82	6.17	7.27	10.46	12.23	17	5	0.12	0.38	0.31	0.91
8.1	10	20	13.96	2.52	9.54	11.82	15.49	19.75	24	4	0.10	0.56	0.23	0.99
8.1	11	20	13.17	2.60	8.10	11.25	15.00	17.67	18	5	0.12	0.39	0.29	0.92
8.1	12	19	14.40	2.91	7.77	12.83	16.37	18.47	20	7	0.12	0.29	0.27	0.78
8.1	1	16	14.36	2.62	9.55	11.92	16.53	18.15	14	6	0.14	0.30	0.38	0.80
8.1	2	16	15.81	3.51	6.53	13.92	18.83	19.99	13	8	0.15	0.23	0.40	0.65
8.1	3	14	13.52	4.62	5.23	8.84	16.72	21.31	19	5	0.14	0.37	0.28	0.91
8.1	4	13	12.16	4.09	7.14	8.02	16.17	17.77	11	7	0.17	0.28	0.49	0.76
8.1	5	15	11.24	3.98	5.40	7.13	14.67	17.20	13	6	0.15	0.32	0.40	0.83
8.1	6	18	11.96	2.98	7.06	8.99	14.31	17.10	15	5	0.13	0.39	0.35	0.92
8.1	7	19	12.77	2.20	9.60	11.50	14.22	17.78	20	4	0.11	0.56	0.26	0.99
8.1	8	17	12.69	2.13	8.76	11.08	14.59	16.31	12	5	0.15	0.38	0.42	0.91
8.1	9	20	13.31	2.87	8.33	10.96	15.85	17.44	13	6	0.15	0.30	0.42	0.80

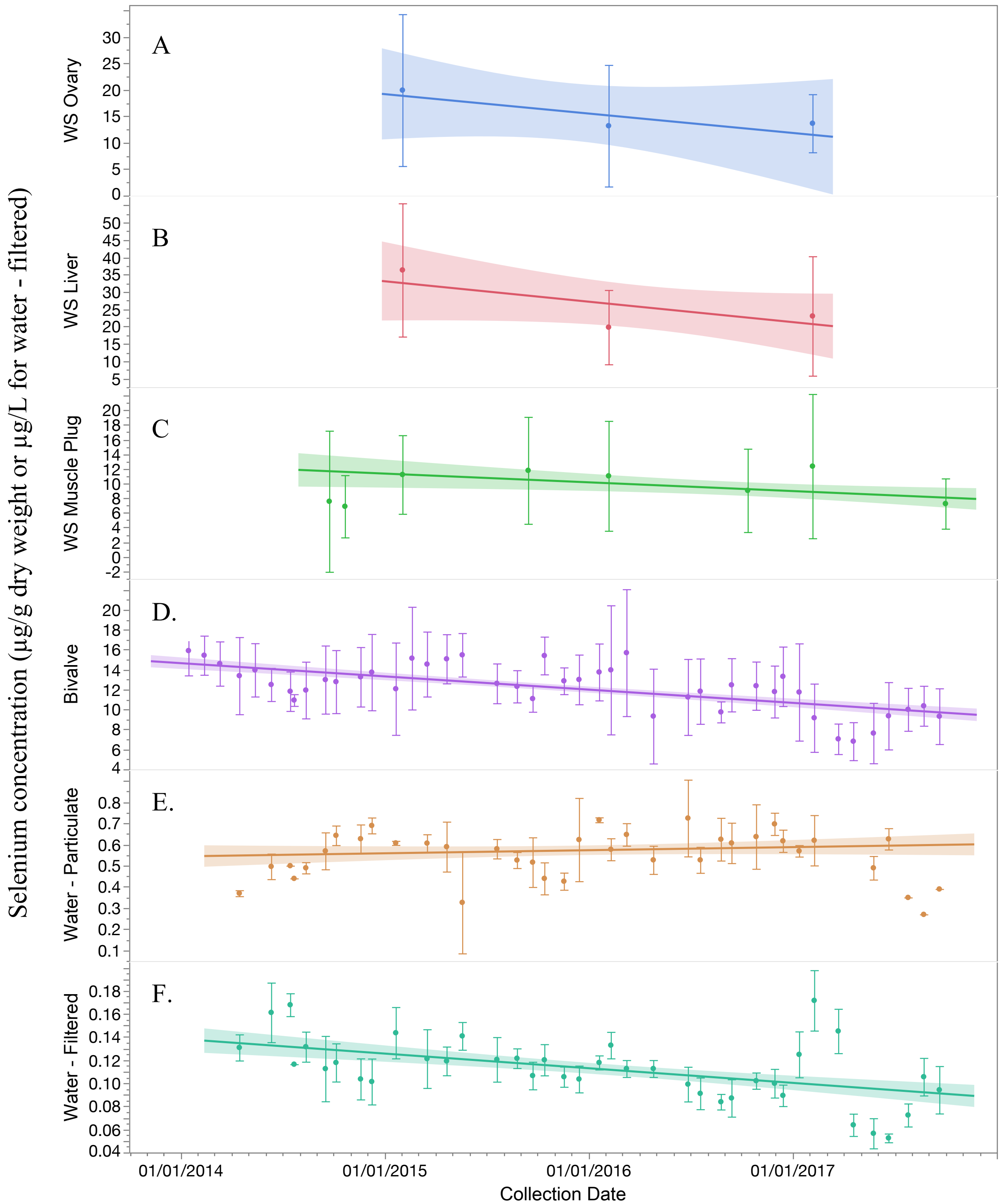

Figure 5-2. Selenium concentration in monitoring matrices from April 2014 through September 2017. A. White sturgeon – Ovary. B. White sturgeon – Liver. C. White sturgeon – muscle plugs. D. Bivalves – Potamocorbula amurensis. E. Water – particulate. F. Water – filtered. Values are means in µg/g dry weight (except filtered water which is in µg/L) ±SD for each sample date, whereby SDs represent replicated sites (water, particulate and bivalves – Stations 2.1, 4.1, 8.1, and 12.5) and individual white sturgeon. Regression lines ±95% confidence bands represent long-term trend through time series.

6 Evaluation of Monitoring Strategies for Sturgeon Selenium

Two white sturgeon selenium monitoring efforts are being conducted in San Francisco Bay. Long-term monitoring is being conducted by the RMP's Status and Trends sport fish monitoring effort; it occurred every three years from 1994 to 2009 and then every five years since 2009, and it is planned to continue on a five-year cycle (i.e., 2014, 2019, etc.). This monitoring includes the analysis of selenium in twelve individual sturgeon caught in North to Lower South San Francisco Bays. On an ongoing basis, only three individual sturgeon will be collected from North Bay every five years.

Since 2014, the RMP has been piloting annual fall muscle plug monitoring through a collaboration with the California Department of Fish and Wildlife (CDFW). Each year, this monitoring has targeted the collection of muscle plugs from 60 sturgeon in the size range of 100-160 cm fork length (the approximate size range of the sturgeon slot limit for fishing in San Francisco Bay, 40-60 inches, equivalent to approximately 115-181 cm total length) sampled in North Bay in September and October. Through this pilot work, the muscle plug sampling approach has been established as the most feasible, cost-effective, and least invasive way to collect muscle tissue samples from a large number of sturgeon for selenium monitoring. The goal of the present analysis is to refine the sampling design for this ongoing monitoring effort.

Data from several prior studies are also available for use in this analysis. Sturgeon muscle selenium monitoring has been conducted periodically since the late 1980s, including the State Water Board's Selenium Verification Study in 1986-1990 (CSWRCB, 1987, 1998, 1989, 1991) and studies by Stewart et al. (2004), and Linares-Casenave et al. (2015a,b), in addition to the RMP Status and Trends sport fish monitoring reports (most recently, Sun et al., 2017), and the 2015-2017 RMP Sturgeon Derby and RMP Muscle Plug reports (Sun et al., 2018a,b in prep). As part of the effort to inform future monitoring efforts and data analysis, recently collected RMP muscle plug data were analyzed in the context of this historically available sturgeon muscle selenium dataset for the Bay (Sun et al., 2018a in prep).

Several factors potentially contributing to sturgeon selenium variability were evaluated to inform efforts to 1) constrain or eliminate variability in the long-term monitoring design, and 2) control for remaining variability in statistical analyses, thereby increasing power for detection of changes in sturgeon selenium concentrations. These included environmental factors that affect dietary selenium, including annual and seasonal hydrology (freshwater inflow from the Delta, assessed using water year type and month of sampling) and foraging location and biological factors (fish length or age, sex, and reproductive stage). Those analyses and their implications for long-term monitoring design are summarized in Table 6-1.

Table 6-1. Summary of key factors influencing selenium accumulation in sturgeon.

Factor	Description
Water Year Type	Data from the 2015-2017 RMP Muscle Plug study suggest that water year type has a significant effect on sturgeon selenium concentrations, matching expectations based on bivalve selenium patterns. This factor introduces significant inter-annual variability into the long-term time series, and should be included as a covariate in future trend analyses.
Season	Historical data are not sufficient to statistically evaluate this factor. Qualitative data analysis and understanding of environmental and physiological processes suggest that higher concentrations should be expected in spring than summer or fall. Long-term monitoring in collaboration with the CDFW sturgeon tagging survey will occur during only one season (fall). The effect of season on sturgeon selenium will not affect the detection of long-term trends based on consistent fall sampling with the CDFW.
Capture (and Presumed Foraging) location	Historical data suggest that higher concentrations are found in North Bay, which is the area of regulatory interest. Long-term monitoring through the CDFW sturgeon tagging survey will occur only in North Bay. This factor will not affect the detection of long-term trends.
Length/ Age	Historical data suggest that juveniles have lower muscle selenium concentrations than adults. No clear size-related trend in selenium concentrations has been found among adults. Future monitoring will focus only on adults (>105 cm total length). Constraining monitoring to adults will reduce variability while focusing on the main population of interest for future monitoring. Further evaluation of the significance of this factor and any potential interaction effects can be evaluated by including this factor as a covariate in future trend analyses.
Sex	Historical data suggest that there is no significant effect of sex on selenium concentrations. Previous sampling included use of blood plasma for sex determination, but collection of blood samples for sex steroid analyses is not necessary during future monitoring.
Reproductive Stage	There are not enough historical data to evaluate this effect, or to include it in the model. USEPA monitoring guidance indicates that monitoring does not need to be designed to target a segment of the population based on reproductive stage. Similarly, previous sampling included use of blood plasma for reproductive stage determination, but collection of blood plasma for those analyses is not necessary during future monitoring.

6.1 Monitoring Program Design Analyses

6.1.1 Linear Regression

The primary means of evaluating the statistical design of the sturgeon monitoring was based on the power of linear regression analyses to detect increases in sturgeon selenium concentrations. The overall goals of the long-term monitoring strategy for sturgeon are to evaluate long-term increases in selenium exposure in the food web and to compare muscle selenium concentrations to the TMDL target of 11.3 µg/g dw (Baginska, 2015). Sturgeon respond to changes in selenium more slowly than bivalves or the water column itself, but they offer the benefit of providing a spatially and temporally integrated representation of selenium exposure. Therefore, sturgeon will be valuable indicators of long-term trends in food web exposure and can be used to compare to the TMDL target for muscle, but will not be a leading indicator of changes in selenium sources.

6.1.2 Method Description

Data collected during the 2015-2017 RMP Muscle Plug study were considered a present-day baseline condition. To evaluate the power to detect future increases in sturgeon selenium concentrations, a synthetic time series was created using the data distribution for adults from the Muscle Plug Study (Table 6-2). Adults were defined as sturgeon with total length > 105 cm, following size-classes established by Linares-Casenave et al. (2015). In the 2017 Muscle Plug Study total lengths were not measured directly, so they were estimated based on fork length, using a fork length-total length regression calculated from all available studies of sturgeon selenium in the Bay that reported both total and fork length (RMP S&T, 2014 [Sun et al., 2017], 2015-2016 Muscle Plug studies [Sun et al., 2018a in prep], Linares-Casenave et al., 2015). The size range targeted by the Muscle Plug studies (100-160 cm fork length, equivalent to approximately 115-181 cm total length) includes only adults, following this classification.

Because sturgeon muscle selenium concentrations are log-normally distributed, the data were log-transformed for simulation and analysis. The approach is briefly described as follows.

- Create trended annual means by beginning with the mean of the log-transformed 2015-2017 RMP muscle plug data, and adding an increase of the mean concentrations of between 1 and 3% per year (in untransformed data).
- Account for inter-annual variability by adjusting the annual means using the standard deviation of the annual means of the log-transformed 2015-2017 RMP muscle plug data. The inter-annual variability (standard deviation of the annual means) will account for the effects of factors such as hydrology, given that 2015-2017 spanned critically dry, dry, and wet water years.
- Account for intra-annual variability by creating data surrounding the annual means using the standard deviation of the entire log-transformed dataset from the 2015-2017 RMP Muscle Plug study.
- Run 10,000 simulations for each scenario and aggregate results to assess detection power.
- Key parameters for the simulations are listed and explained in Table 6-2.

Table 6-2. Key simulation parameters.

Parameter	Value	Value Description
Fixed		
Mean (µg/g dw)	2.07	Mean of the log-transformed 2015-2017 muscle plug data
Intra-annual variation	0.57	Standard deviation of the log-transformed 2015-2017 RMP muscle plug data
Inter-annual variation	0.22	Standard deviation of the annual means of the log-transformed 2015-2017 RMP muscle plug data
Scenarios		
Inter-annual variation	Two levels	Normal: 0.22 (see above) Reduced: 0.11 (half of normal)
Sampling frequency	Four levels	Every 1, 2, 3, or 5 years
Sample numbers	Two levels	30 or 60 samples per year
Increases in sturgeon selenium concentrations	Three levels	1%, 2%, or 3% increase per year

6.1.3 Results

Results for the power to detect an increase in sturgeon muscle selenium concentrations over 10 or 20 years are shown in Table 6-3. Monitoring frequencies were varied from annual to every 2, 3, or 5 years. Three trend scenarios were run, including increases in selenium concentrations of 1%, 2%, or 3% per year.

Over a 20-year period, trends in the range of 2 or 3% increase per year are likely to be detectable with feasible monitoring designs. Assuming inter-annual variability at the level measured in 2015-2017, the detection of a 2% increase per year is feasible with annual monitoring of 30 samples per year and with both annual or biennial monitoring of 60 samples per year (power > 0.80). The detection power increases slightly for the 20-year period under the hypothetical condition of reduced inter-annual variability so that a 2% increase per year would be detectable with 30 samples in biennial monitoring. Inter-annual variability calculated from the 2015-2017 period may be higher than what might occur during a longer time span, assuming a significant proportion of this variability is driven by hydrology: both critically dry and wet water years were experienced during those 3 years. At a lower level of inter-annual variability, a 2% increasing trend per year would almost be detectable with power greater than 0.8 with triennial monitoring of 60 samples per year.

A 3% per year increase over 20 years would be detectable with substantially less monitoring – either biennial monitoring of 30 samples per year, or as infrequently as monitoring 60 samples every 5 years. Assuming a lower inter-annual variability over the long term, such a trend would also be detectable with 30 samples collected only triennially (power > 0.80), or almost detectable if collected once every 5 years (power= 0.79).

Over a 10-year period, there would be low power to detect an increase. The only scenario with power greater than 0.8 was with a trend of 3% increase per year, and annual sampling of 60 plugs with the lower degree of inter-annual variation.

Table 6-3. Power to detect significant increases in sturgeon selenium concentrations for various design scenarios. Power equal to or greater than 0.80 is shown in bold.

	20 Years			10 Years		
Inter-annual variation 0.22	Trend: 3% per year	30 plugs	60 plugs	Trend: 3% per year	30 plugs	60 plugs
	1 year	**0.99**	**1.00**	1 year	0.60	0.74
	2 years	**0.91**	**0.95**	2 years	0.51	0.63
	3 years	0.75	**0.84**	3 years	0.43	0.58
	5 years	0.71	**0.80**	5 years	0.43	0.57
Inter-annual variation 0.22	Trend: 2% per year	30 plugs	60 plugs	Trend: 2% per year	30 plugs	60 plugs
	1 year	**0.91**	**0.95**	1 year	0.50	0.60
	2 years	0.73	**0.83**	2 years	0.43	0.57
	3 years	0.59	0.70	3 years	0.42	0.55
	5 years	0.54	0.67	5 years	0.41	0.54
Inter-annual variation 0.22	Trend: 1% per year	30 plugs	60 plugs	Trend: 1% per year	30 plugs	60 plugs
	1 year	0.59	0.72	1 year	0.36	0.50
	2 years	0.49	0.64	2 years	0.39	0.52
	3 years	0.43	0.57	3 years	0.38	0.52
	5 years	0.42	0.56	5 years	0.40	0.53
Inter-annual variation 0.11	Trend: 3% per year	30 plugs	60 plugs	Trend: 3% per year	30 plugs	60 plugs
	1 year	**1.00**	**1.00**	1 year	0.67	**0.85**
	2 years	**0.98**	**1.00**	2 years	0.45	0.63
	3 years	**0.86**	**0.96**	3 years	0.31	0.45
	5 years	0.79	**0.92**	5 years	0.26	0.40
Inter-annual variation 0.11	Trend: 2% per year	30 plugs	60 plugs	Trend: 2% per year	30 plugs	60 plugs
	1 year	**0.98**	**1.00**	1 year	0.43	0.61
	2 years	**0.85**	**0.95**	2 years	0.31	0.46
	3 years	0.61	0.77	3 years	0.24	0.36
	5 years	0.54	0.73	5 years	0.21	0.34
Inter-annual variation 0.11	Trend: 1% per year	30 plugs	60 plugs	Trend: 1% per year	30 plugs	60 plugs
	1 year	0.63	**0.82**	1 year	0.23	0.34
	2 years	0.44	0.61	2 years	0.20	0.32
	3 years	0.31	0.46	3 years	0.18	0.29
	5 years	0.28	0.41	5 years	0.18	0.28

Figures 6.1 and 6.2 show examples of simulated data and means for different scenarios.

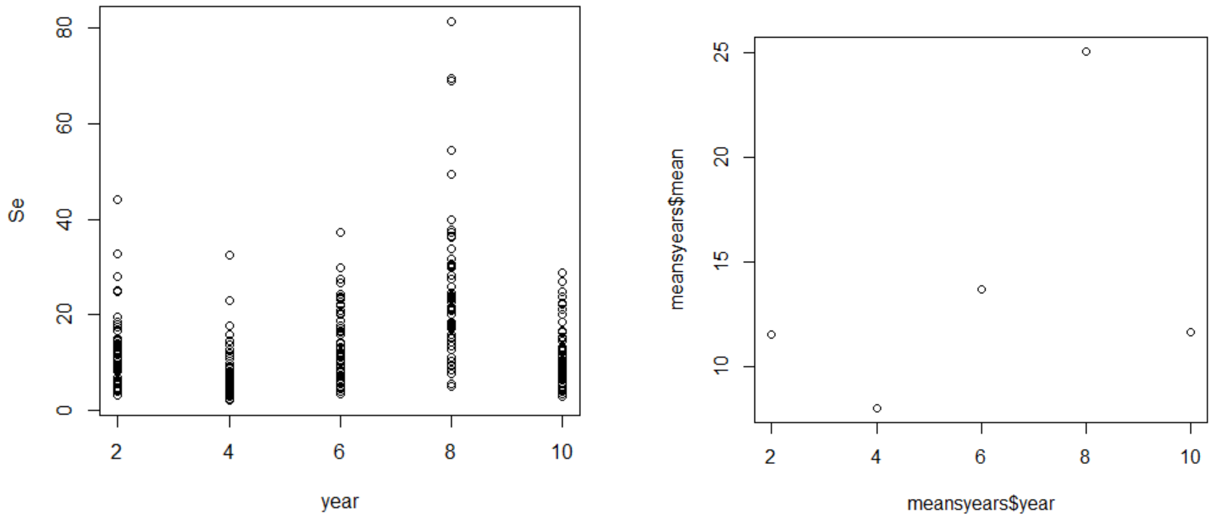

Figure 6-1. Simulated data and means for a scenario with 60 plugs sampled every 2 years over a 10-year period (normal inter-annual variability, 3% per year increasing trend).

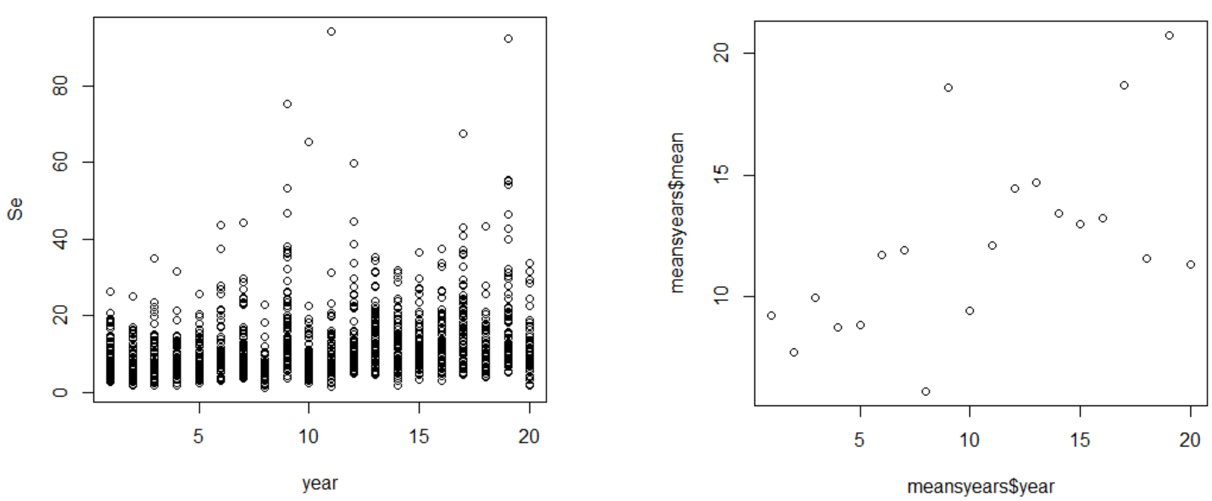

Figure 6-2. Simulated data and means for a scenario with 60 plugs sampled every year over a 20-year period (normal inter-annual variability; 3% per year increasing trend).

Ultimately, as additional data are collected to better characterize the size of the effect of water year type (or flow) on sturgeon selenium concentrations, it may be possible to normalize the raw data by hydrologic condition. Reducing inter-annual variability would increase power as suggested by the lower-variance scenarios.

6.2 Implications for Future Monitoring

The results of the design analyses highlight the high level of variability in sturgeon muscle selenium data and consequently the time needed to be able to detect long-term trends of increasing selenium concentrations. Assuming a lower level of inter-annual variability compared to 2015-2017, a 3% increase in the mean concentration per year would be detectable within 10 years with annual monitoring of 60 samples per year. Detecting lower rates of change or detecting this change under higher inter-annual variability conditions would require monitoring over a longer time frame. Detecting a 1% per year trend would likely require monitoring for more than 20 years.

Trends of 2 or 3% per year are likely to be detectable within 20 years. In general, increasing the sample number from 30 to 60 substantially increases detection power, particularly for scenarios with lower overall concentration increases. It should be noted that the current simulation of log-transformed selenium data can result in the occasional creation of high concentration data points that may not reflect potential real conditions (concentrations at 50-150 ug/g dw, or occasionally higher), particularly during longer simulations with a greater number of data points; this may inflate the true detection power. However, in general the monetary cost of increasing the sample numbers from 30 to 60 is relatively low (estimated laboratory analysis cost of $5,700 per 30 plugs). Therefore, future collection of at least 60 samples each sampling event is recommended.

Assuming the analysis of 60 samples per sampling event over a 20-year period, a minimum of biennial monitoring would be needed to detect trends at a 2% rate of increase per year –about a 50% increase over the entire monitoring period. Less frequent monitoring would be needed to detect larger increases in selenium concentrations, but a 3% rate of increase per year is considerable, and may be unlikely to occur. Annual monitoring should be considered if detection of smaller trends, or trends over a shorter period, is desired.

6.3 Proposed Monitoring Framework for Sturgeon

6.3.1 Goals of the Monitoring Program

Monitoring of sturgeon muscle selenium has the following goals:

- Tracking attainment/impairment (management question 1)
- Confirmation of clam and water trends (management question 2)
- Long-term trend evaluation (not early warning) (management questions 1 and)

White sturgeon and green sturgeon are among the sensitive species that the TMDL aims to protect (Baginska, 2015), so selenium concentrations in sturgeon tissue are the ultimate index of impairment (management question 1). Monitoring sturgeon at a greater frequency than once a year is not logistically feasible, and the existing dataset to design such monitoring is limited (the "normal range" is not well-established), so sturgeon monitoring is not well-suited to be an early indicator of change in selenium exposure in the North Bay. However, sturgeon monitoring is needed to confirm that the early warning indicators are signaling a change that propagates to the species of concern (management question 2). Sturgeon monitoring also has value in tracking trends on a long-term time scale.

6.3.2 Sampling Design

- Parameters: Total selenium in muscle plugs; C, N, S isotopes in muscle plugs
- Location: Suisun, San Pablo Bays (depending on where CDFW tagging occurs)
- Sampling frequency: Biennial
- Replication: Plugs from 60 sturgeon per event
- Analytical approach: ID-HG-ICP-MS

6.3.3 Cost of the Monitoring Program

- Average annual cost of $23,000 per year.

6.3.4 Monitoring Program Performance

- 83% power to detect 2% per year increase in 20 years with biennial sampling
- Sturgeon monitoring is valuable in tracking North Bay attainment/impairment and whether change is occurring

6.3.5 Statistical Test Methods and Plans for Analysis of the Data

- Linear regression of long-term time series

7 Summary

Table 7-1 provides a summary of information presented in Sections 4 – 6 and implementation details for moving forward with the development of an enhanced selenium monitoring program in NSFB and the Delta. A shared goal of the monitoring design for the three environmental indicators of change is to establish an integrated monitoring plan that can detect increases in selenium concentrations in the Bay and Delta that warrant prompt management responses to signs of increasing risk of impairment. These proposed designs are intended to provide the information required to address the following management questions identified by the RMP's Selenium Workgroup:

1. Are the beneficial uses of North San Francisco Bay impaired by selenium?
2. Are changes occurring in selenium concentrations that warrant changes in management actions?
3. Will proposed changes in water flows and/or selenium loads in the Bay or upstream cause impairment in the North Bay?

One of the main contributions of this design effort is the evaluation of expected monitoring program performance. Each of the technical sections evaluated the statistical power (defined as the probability that the dataset will be sufficiently sensitive to detect a change or trend of a specified magnitude) that will be achieved. Overall, the findings show that the implementation and continuation of long-term monitoring programs are required to identify changes from established baselines and to distinguish deviations from the effects of natural variability. The results from the long-term sampling of bivalve tissue to date, described in Section 5, demonstrate the potential return on the monitoring investment. The clam tissue dataset that was generated over the past 22 years provides the ability to characterize the effects of seasonality and to distinguish between the effects of annual variation in freshwater flow conditions and changes in system loading.

Much of the current understanding of the mechanisms affecting selenium concentrations in NSFB and the Delta is based on modeling results. The linkage of water-column, clam and sturgeon sampling and the continued monitoring of selenium across all water-year types will make an important contribution to the evaluation of the relative importance of hydrologic variability in the system, selenium concentrations and loading levels, and operational changes or new infrastructure on bioaccumulation. Finally, while the data generated from the continued monitoring will greatly benefit the understanding of the behavior of selenium in the system, the information gained on the effects of hydrology and human-induced changes to the system would also contribute toward understanding the behavior of other contaminants in the system.

Table 7-1. Framework for Evaluation of Environmental Indicators.

Table 7-1. Framework for Evaluation of Environmental Indicators			
Overall Se Monitoring Program Goals – Develop sampling strategies for the three identified indicators of change and establish a monitoring plan that can detect increases in selenium concentrations in the Bay and Delta that warrant prompt management responses to signs of increasing risk of impairment. Address management questions: 1. Are the beneficial uses of North San Francisco Bay impaired by selenium? 2. Are changes occurring in selenium concentrations that warrant changes in management actions? 3. Will proposed changes in water flows and/or selenium loads in the Bay or upstream cause impairment in the North Bay?			
	Program Elements		
	Water Column Dissolved and Particulate Selenium	**Bivalve Tissue**	**Sturgeon Tissue**
1. Overview of the Monitoring Element	• The NSFB Se TMDL and USEPA's proposed aquatic life and aquatic-dependent wildlife criteria for the Bay and Delta established targets for dissolved water column Se. • Dissolved and particulate Se concentrations provide key inputs to bioaccumulation models. • No systematic water-column Se sampling program in Delta or Bay.	• The bivalve-tissue data set provides a 22-year time series spanning a broad range of freshwater inflows and physicochemical conditions. • These data define the expected range of tissue concentrations from which change can be identified.	• The NSFB Se TMDL and USEPA's proposed aquatic life and aquatic-dependent wildlife criteria for the Bay and Delta established targets for Se in sturgeon muscle. • RMP pilot study conducted to determine the most feasible, cost-effective, and least invasive way to collect muscle tissue samples from a large number of samples.
2. Sampling Design	• Total, dissolved, particulate Se, TSM, TOC, Chl a • Eight stations; Suisun and San Pablo Bays, Delta • Single monthly grab samples	• Total Se in bivalve soft tissue; C, N, S isotopes • USGS Stations 4.1 and 8.1 • Frequency: 6 months • 5 composites, bivalve sizes from 9 16 mm. • Analytical approach: ID-HG-ICP-MS	• Total Se, C, N, S isotopes in muscle plugs • Suisun, San Pablo Bays • Plugs from 60 sturgeon per event • Biennial sampling
3. Total Cost of the Monitoring Program	• Average annual cost of $93K in combination with the bivalve monitoring.	• Average annual cost of $93K in combination with the water monitoring.	• Average annual cost of $23K
4. Statistical Testing Methods and Plans for Analysis of the Data	• Change Point Detection Method • Normal range methods	• Normal range methods	• Linear regression of long-term time series
5. Monitoring Program Performance	• Greater than 80% likelihood of detecting a 20% change with detection times within a 10 yr sampling period.	• Greater than 90% likelihood of detecting exceedance of the 97.5th quantile of the distribution of the existing data.	• Greater than 80% power to detect 2% per yr increase in 20 yr with biennial sampling.
6. Added Value	• Contribute to understanding of the role of hydrologic variability and modifications in the system to observed changes in bioaccumulation. • Demonstrate effectiveness of improved temporal trend monitoring methods.		• Ultimate indicator for impairment assessment.

8 References

Arciszewski T.J., Munkittrick K.R. 2015. Development of an adaptive monitoring framework for long-term programs: An example using indicators of fish health. *Integrated Environmental Assessment and Management* 11: 701-718.

Arciszewski T.J., Munkittrick K.R., Scrimgeour G.J., Dubé M.G., Wrona F.J., Hazewinkel R.R. 2017. Using adaptive processes and adverse outcome pathways to develop meaningful, robust, and actionable environmental monitoring programs. *Integrated Environmental Assessment and Management* 13: 877-891.

Baginska, B. 2015. Total Maximum Daily Load Selenium in North San Francisco Bay, Staff Report for Proposed Basin Plan Amendment. On the Internet at: http://www.waterboards.ca.gov/sanfranciscobay/board_info/agendas/2015/November/6_appendix_c.pdf

California State Water Resources Control Board (CSWRCB). 1987. Selenium verification study, 1986. Sacramento (CA): California State Water Resources Control Board. 79 p + 9 appendices.

California State Water Resources Control Board (CSWRCB). 1988. Selenium verification study, 1986–1987. Sacramento (CA): California State Water Resources Control Board. 60 p + 8 appendices.

California State Water Resources Control Board (CSWRCB). 1989. Selenium verification study, 1987–88. Sacramento (CA): California State Water Resources Control Board. 81 p + 11 appendices

California State Water Resources Control Board (CSWRCB). 1991. Selenium verification study, 1988–1990. Sacramento (CA): California State Water Resources Control Board. 79 p + 9 appendices.

California Waterfix, 2018. Fixing California's Water System. [ONLINE] Available at: https://www.californiawaterfix.com/. [Accessed 28 November 2018].

Cutter, G.A., and L.S. Cutter. 2004. Selenium biogeochemistry in the San Francisco Bay estuary: Changes in water column behavior. *Estuarine, Coastal and Shelf Science* 61: 463–476.

Doblin, M.A., S.B. Baines, L.S. Cutter, Cutter, G.A. 2006. Sources and biogeochemical cycling of particulate selenium in the San Francisco Bay estuary. *Estuarine, Coastal and Shelf Science* 67:681–694.

Gitzen, R.A., J.J. Millspaugh, A.B. Cooper, Licht, D.S. eds., 2012. Design and analysis of long-term ecological monitoring studies. Cambridge University Press.

Kilgour B.W., Somers K.M., Barrett T.J., Munkittrick K.R., Francis A.P. 2017. Testing against "normal" with environmental data. *Integrated Environmental Assessment and Management* 13: 188-197.

Kleckner, A.E., Kakouros E., Stewart A.R. 2017. A practical method for the determination of total selenium in environmental samples using Isotope Dilution-Hydride Generation-Inductively Coupled Plasma-Mass Spectrometry. *Limnology and Oceanography: Methods* 15: 363-371.

Linares-Casenave, J., R. Linville, J.P. Van Eenennaam, J.B. Muguet, Doroshov, S.I. 2015a. Selenium tissue burden compartmentalization in resident white sturgeon (*Acipenser transmontanus*) of the San Francisco Bay Delta Estuary. *Environmental Toxicology and Chemistry* 34: 152-160.

Linares-Casenave, J., R. Linville, J.P. Van Eenennaam, J.B. Muguet, Doroshov, S.I. 2015b. Selenium tissue burden compartmentalization in resident white sturgeon (*Acipenser transmontanus*) of the San Francisco Bay Delta Estuary: Corrigendum. *Environmental Toxicology and Chemistry* 34: 943.

Linville R.G., Luoma S.N., Cutter L., Cutter G.A. 2002. Increased selenium threat as a result of invasion of the exotic bivalve *Potamocorbula amurensis* into the San Francisco Bay-Delta. *Aquatic Toxicology* 57: 51-64.

Ohlendorf H.M., Lowe R.W., Harvey T.E., Kelly P.R. 1986. Selenium and heavy metals in San Francisco Bay diving ducks. *Journal of Wildlife Management* 50: 64-71.

Ohlendorf H.M., Marois K.C., Lowe R.W., Harvey T.E., Kelly P.R. 1991. Trace elements and organochlorines in surf scoters from San Francisco Bay, 1985. *Environmental Monitoring and Assessment* 18: 105-122.

Ross, G.J., 2015. Parametric and nonparametric sequential change detection in R: The cpm package. *Journal of Statistical Software* 66(3): 1-20.

SAS Institute Inc., 2018. JMP® 14 Fitting Linear Models. SAS Institute Inc, Cary, NC.

Stewart A.R., Luoma S.N., Schlekat C.E., Doblin M.A., Hieb K.A. 2004. Food Web Pathway Determines How Selenium Affects Aquatic Ecosystems: A San Francisco Bay Case Study. *Environmental Science & Technology* 38: 4519-4526.

Stewart, A.R., S.N. Luoma, K.A. Elrick, J.L. Carter, van der Wegen, M. 2013. Influence of estuarine processes of spatiotemporal variation in bioavailable selenium. *Marine Ecology Progress Series* 492: 41-46.

Stewart A.R., Jongebloed U., Kleckner A.E.. in prep. Unprecedented drought results in elevated and persistent bioavailable selenium in the San Francisco Estuary.

Sun, J., J.A. Davis, S.N. Bezalel, J.R.M. Ross, A. Wong, R. Fairey, A. Bonnema, D.B. Crane, R. Grace, R. Mayfield, and J. Hobbs. 2017. Contaminant Concentrations in Sport Fish from San Francisco Bay. San Francisco Estuary Institute-Aquatic Science Center, Richmond, CA.

Sun, J., J.A. Davis, and A.R. Stewart. 2018a, in prep. Selenium in Muscle Plugs of White Sturgeon from North San Francisco Bay, 2015-2017. San Francisco Estuary Institute-Aquatic Science Center, Richmond, CA.

Sun, J., J.A. Davis, A.R. Stewart, and V. Palace. 2018b, in prep. Selenium in White Sturgeon Tissues: 2015-2017 Sturgeon Derby. San Francisco Estuary Institute-Aquatic Science Center, Richmond, CA.

Tetra Tech, 2012. North San Francisco Bay Selenium Characterization Study. Final Report. http://www.waterboards.ca.gov/sanfranciscobay/water_issues/programs/TMDLs/northsfbayselenium/DSM2%20Selenium%20Loads%20from%20the%20Delta%205-1-2014.pdf

Tetra Tech, 2017a. Water Column Selenium Concentrations in the San Francisco Bay-Delta: Recent Data and Recommendations for Future Monitoring. August 2017. SFEI Contribution #836.

Tetra Tech, 2017b. Proposed Plan for Enhanced Selenium Monitoring in the Delta to Track Future Loading Changes Associated with the California WaterFix Project. Prepared by Tetra Tech for the Western States Petroleum Association. December 1, 2017.

U.S. Environmental Protection Agency (USEPA). 2016. Proposed Rule, Water Quality Standards; Establishment of Revised Numeric Criteria for Selenium for the San Francisco Bay and Delta, State of California. On the Internet at: https://www.federalregister.gov/documents/2016/07/15/2016-16266/water-quality-standards-establishment-of-revised-numeric-criteria-for-selenium-for-the-san-francisco