Final small fish spatial survey design
Ben Greenfield, SFEI
March 11, 2008

Summary

This document recommends a specific sampling design for evaluating spatial patterns in mercury (Hg) in biosentinel test organisms in San Francisco Bay. It has formed the basis for work underway to select sampling stations for the small fish spatial program. Additional study components (seasonal and interannual variation) are not addressed in this document.

Table 1. Overview of the small fish spatial survey design.

Hydrology:	*Open Water*	*Embayment*	*Total Number*
Station Type			
High Hg Wetland	NA	NA	5-7/yr
Low Hg Wetland	NA	NA	5-7/yr
Hg Source Area	5-7/yr	5-7/yr	10-14/yr
Other	5-7/yr	5-7/yr	10-14/yr
Total number sites			30 – 42/yr

The primary features of the design are:

- A stratified random design
- Stratified by site hydrology (open water versus embayment), as well as site type
- 1/3 wetlands; 1/3 Hg source sites; 1/3 other
- Will be repeated with new sites annually, to increase statistical power and coverage.

Background and hypotheses

The goal of the project is to determine hotspots of methylmercury bioavailability by monitoring mercury concentrations in small fish and sediments. For a variety of reasons, fish and sediments are considered to be appropriate monitoring tools. Monitoring will also be performed on Hg isotopes in fish and with Diffusive Gradient in Thinfilm (DGT) Devices, in coordination with this program.

Substantial published and unpublished data on the system lead to a number of predictions and hypotheses regarding the conditions in which methylmercury (hereafter, mercury) will be higher. For example, previous work indicates that concentrations tend to be higher near the shoreline of the system (including enclosed subembayments such as salt ponds and wetlands), and lower in open waters. Therefore, this study focuses on the statistical population of nearshore areas, and excludes deep water offshore areas of the Bay. Spatial autocorrelation of data points is also expected (there are several reasons for this). Therefore, we are using the Generalized Random Tessellation Stratified (GRTS) Spatially-Balanced Survey Designs GRTS to obtain a spatially balanced random sample.

The spatial survey is interested in answering two basic questions:

1. What factors (i.e., site characteristics) appear to be important for causing increased mercury concentrations in Bay biota?
2. Where are the highest mercury concentrations found in the nearshore portions of the system?

Question 2 should be answerable as data are obtained in any stratified randomized design. The project will also test the following hypotheses regarding the factors that are likely to be most important (Question 1):

1. mercury concentrations will be elevated at wetlands
2. mercury concentrations will be most elevated in specific types of wetlands (wetlands that have intermittent wetting and drying).
3. Concentrations will increase along a gradient from north to south.
4. Concentrations will be higher in fully or partially enclosed areas. This would include subembayments, natural or man-made channels, or estuarine creeks draining into the Bay. These areas would tend to have low hydraulic mixing of subtidal water (i.e., locations with low water turnover rate). Hereafter, these types of areas will be referred to as "embayments".
5. Concentrations will be higher near mercury source areas. Unfortunately, there is a wide range of different types of potential source areas. Table 2 summarizes the types of source areas, and estimates of the number of sites (if known) that exist in the population.
6. Concentrations will be lower at sites that don't contain any of these attributes. For example, Bay shore lines not near particular sources and adjacent to well mixed open waters.

Table 2. Types of mercury source areas. Note that these are the third row from Table 1.

Source type	**Total number of sites in the universe***
Creek draining urban or industrial area	30 locations
Creek draining watershed with hydrologically connected mercury mines	4 locations
Treated wastewater outfall flowing into low turnover area	11 locations
Areas with high sediment Hg deposits due to historic industrial activity	Around 5% of Bay shoreline[1]

*all of these values are estimates-we would have to do research to identify all the sites

<u>Design constraints</u>

The total pool of stations that we will be able to sample is approximately 30 - 42 per year[2]. For this description of sampling design, we will assume that 42 stations are available – but the actual number will depend on the budget. The survey is scheduled to be performed for three years- however we may be asked to change the design after the first or second year. At each station, we will be targeting two fish species: topsmelt and Mississippi silverside. At some stations, only one

---

[1] Fifteen stations were identified with sediment Hg > 0.7 ppm in the Bay Protection and Toxic Cleanup Program
[2] Pending budget analysis and need for matching funds to other studies

species will be successfully captured. We expect to collect about four composite samples per fish species. To the extent possible, fish will be sampled within a fixed body size range, with each composite similar in size. Total body length will have to be included as a predictor variable in any statistical models that are developed.

Design

Based on all of this, the following design structure will be used (see also Table 1).

Six site categories would be sampled - each with equal frequency:

1. Wetlands expected to be high in mercury (intermittent wetting and drying events)
2. Wetlands not expected to be high in mercury
3. Mercury source areas (see Table 2) in embayments, and other enclosed locations
4. Mercury source areas (Table 2) in open water locations
5. Other sites in embayment locations
6. Other sites in open water locations

If we are able to sample 42 sites per year, this would result in 7 sites from each of these categories per year.

All randomization will be performed with the GRTS protocol, based on GIS coverages from the different categories. The following procedures :

- mercury source areas would be selected from among a large pool of individual sites taken from Table 2. Because there are many types of source areas, an effort would be made to sample each of them with equal probability. This limits our ability to distinguish among the different types of source areas as potential mercury hotspots.
- embayment vs. open water are developed based on best professional judgment over a GIS
- wetland sampling selection has already been performed based on the two wetland categories using a GRTS scheme
- "other sites" would be selected from a GIS layer of the remaining types of Bay shore locations.

One concern that emerges no matter what we do is that some of the sites will invariably be somewhat of a mixture of the different attributes above. For example wetlands associated with industrial areas, or areas adjacent to several of the sources listed in Table 2. This concern will be dealt with in several ways: A. Categorizing sites based on a rank-scale. B. Switching sites among categories, post hoc, if necessary. B. Performing analyses based on continuous variables, such as GIS.

To summarize attributes of the planned data structure (Table 3):

- six categories of station - equal sampling frequency in each category
- two fish species targeted at each station; at some stations only one species is likely to be present;
- fish and sediment isotopes and DGT samples performed at selected stations
- four replicate samples from each fish species

- a randomized stratified design (rather than a fixed design)
- potential for spatial autocorrelation as well as North versus South gradients
- incorporating continuous variables (for example, fish body length, sediment methylmercury, year, and GIS parameters) into the analysis
- the risk that some of the samples, although chosen to represent one category, may actually fit into multiple categories

Table 3. Graphical summary of key attributes of proposed data structure. Response variable is Hg in fish. Notes: This is for one year – design is intended for multiple years. Covariates of concern are a partial list.

	Categories	Stations/ Category	Substations/ Station	Species/ Substation	Samples/ Species
Number	6	7	2 (wetlands) 1 (others)	1-2	4
Type	Fixed	Random	Fixed	Fixed	Random
Total N	6	42	56	100	400
Covariates		Latitude GIS traits Sediment Hg		Body length	

<u>Coordination with Hg Isotopes</u>

A targeted subset of the stations will be selected for sampling Hg isotopes in fish and sediments. A substantial portion of the Hg isotope program will focus on Hg source areas from Table 2. This is based on extensive discussion with Joel Blum during his recent visit to SFEI. The purpose of this focus will be to determine whether fish and sediments from these locations represent distinct isotope signatures associated with specific sources (e.g., legacy Hg mining vs. urban runoff). Separate samples will be chosen to from the ambient sites, with an emphasis on locations distant from particular sources. This is intended to determine whether general ambient bay fish and sediments indicate substantial influence of any particular sources.

<u>Questions regarding design and analysis</u>

<u>Question 1:</u> Is this an appropriate and defensible design to address the questions? If not, what changes need to be made?

<u>Answer (Dr. Trent McDonald, peer review statistician[3]):</u>

The design is appropriate for testing the hypotheses of different potential effects. Because the entire population of Bay samples is sampled (and this is a requirement), the design will be predictive for other stations outside the sample area.

[3] Biometrician/Project Manager, Western EcoSystems Technology, Inc.

It will be possible to characterize overall Bay-wide ambient conditions using this design. However, this will be non-trivial because the relative area of the different categories will vary widely (its essentially the same issue as the RMP S&T program). So the design is not optimized to characterize ambient conditions.

It will be fine to continue the sampling approach, selecting 30 - 42 new stations without replacement, over several years. This will be optimal for characterizing spatial patterns and testing hypotheses. Such a design is weak for trend detection, but trend detection will be accomplished via the separate fixed station design.

<u>Question 2</u> What will be an appropriate statistical approach (or set of approaches) to analyze these data?

<u>Answer (Trent McDonald):</u> If there is no spatial autocorrelation, this will be interpretable using standard General Linear Models (e.g., Proc GLM, in SAS). There will be several categorical variables, with some variables nested within other variables (Table 2). Continuous variables that are measured may also be incorporated. If there is spatial autocorrelation, the assumptions of the model (independence) must be relaxed by incorporating the autocorrelation into the model. This can be performed in SAS using Proc MIXED.