Briefing for OMSAP workshop on ambient monitoring revisions
July 24, 2003

Massachusetts Water Resources Authority

Environmental Quality Department

ENQUAD Report ms-086

Citation: MWRA. 2003. **Briefing for OMSAP workshop on ambient monitoring revisions, July 24, 2003**. Boston: Massachusetts Water Resources Authority. Report ENQUAD ms-086. 26 p.

Information Briefing to Outfall Monitoring Science Advisory Panel

To: Dr. Andrew Solow, Chair OMSAP
From: Dr. Andrea Rex, Director, Environmental Quality, MWRA
Cc: OMSAP, IAAC, PIAC
Date: July 17, 2003
Subject: Briefing materials for review of sediment nutrient flux and infaunal community monitoring design

==

MWRA is in the process of reviewing its outfall monitoring program, using the process described in its NPDES discharge permit. As the Outfall Monitoring Science Advisory Panel requested, a workshop has been scheduled for July 24, 2003 to review the Outfall Ambient Monitoring Plan. The monitoring areas for review include sediment nutrient flux and benthic infaunal community. The attached briefing materials comprise three parts: Introduction, Nutrient Flux study, and Sediment Infauna.

MWRA is proposing to change the design of the special nutrient flux study by discontinuing analyzing for urea flux and pore water analyses. MWRA also plans to update the methodology for nutrient flux analyses.

MWRA is proposing a more efficient design for benthic infaunal sampling, and presents statistical analyses supporting this change in design.

TABLE OF CONTENTS

TABLES AND FIGURES

1. INTRODUCTION

Since 1985, MWRA has worked to minimize the effects of wastewater discharge on the marine environment by ending the discharge of sludge and inadequately treated effluent into Boston Harbor, reducing pollutants at their source, improving wastewater treatment to modern standards, and providing increased dilution. Concerns about potential effects of moving the effluent outfall from the harbor to Massachusetts Bay have been recognized by MWRA and by the joint permit for the outfall issued by the U.S. Environmental Protection Agency (EPA) and the Massachusetts Department of Environmental Protection (MADEP).

MWRA is committed to ensuring good treatment and conducting monitoring necessary to ensure that environmental impact of the discharge is minimal. This commitment is formalized in MWRA's discharge permit, which requires MWRA to monitor the effluent and the ambient receiving waters for compliance with permit limits and in accordance with the monitoring plan (MWRA 1991, 1997a). EPA and MADEP have established an independent panel of scientists, the Outfall Monitoring Science Advisory Panel (OMSAP), to review monitoring data and provide advice on key scientific issues related to the permit. The monitoring plan can be modified, under OMSAP's guidance, to incorporate new scientific information and improved understanding resulting from the monitoring.

In the fiscal year just ended, MWRA spent approximately $5 million on effluent monitoring and environmental studies of Massachusetts Bay. Since 1990, MWRA has invested approximately $35,000,000 in environmental monitoring and modeling.

Table 1-1 Breakdown of July 2002-June 2003 permit-required monitoring costs by project

Project area	MWRA cost	Cost-share
Effluent	$330,000	$79,000
Outfall	$3,702,000	$895,000
water column	1,835,000	595,000
benthos	1,155,000	300,000
fish/shellfish	368,000	
pathogens	344,000	
Model	$136,000	$52,000
Permit reporting and management	$900,000	
Total	$5,068,000	$1,026,000

MWRA's outfall monitoring program (MWRA, 1991, 1997a) was originally designed to determine the effects on Massachusetts Bay of five years of discharge of primary-treated effluent followed by continued discharge of secondary-treated effluent. The expected contaminant loads from the discharge (EPA, 1988) were based on very imprecise estimates. Because of concern about the effects of a primary treated discharge on dissolved oxygen, organic loading to the sea floor, and accumulation of toxic contaminants, the monitoring program was quite comprehensive. Due to outfall construction delays, the secondary treatment plant was completed before offshore discharge started. In addition, effluent flow is lower than had been estimated, meaning more of the flow receives secondary treatment. Finally, measured toxic contaminant concentrations are lower than had been assumed in outfall siting studies even for full secondary treatment.

There are now more than two years of post-discharge monitoring to compare with baseline conditions; monitoring results to date document minimal environmental effect. In addition, the nearly nine years of baseline monitoring provide abundant data to use in evaluating the effectiveness and efficiency of the sampling design. Thus it is appropriate to revisit the monitoring program, as recommended by the National Research Council (NRC, 1990), and refocus it on the potential for long-term chronic effects. Ongoing effluent monitoring will remain at the core of the monitoring program.

A workshop held in March-April 2003 reviewed the fish and shellfish and sediment contaminant monitoring designs (MWRA 2003a.) A workshop held in June, 2003 reviewed water column monitoring (MWRA 2003b.) A comprehensive discussion of effluent monitoring and results took place at both workshops (MWRA 2003a, b) and thus is not included in this briefing.

The monitoring task areas under discussion at the July 24, 2003 OMSAP workshop include soft-bottom community monitoring and the benthic nutrient flux special study. For each area, this briefing package includes a brief summary of the monitoring approach, a description of key results, and MWRA recommendations for monitoring plan modifications and rationale for the changes.

Recommended changes to the soft-bottom community monitoring are reductions in number of stations sampled each year. The intensive monitoring and long-term data set has generated a large database, enabling a detailed statistical evaluation of alternative sampling designs. These changes complement changes in sediment chemistry monitoring discussed at the March/April and June 2003 workshops. Recommended changes to the benthic nutrient flux study include dropping less informative urea and porewater measurements and developing an improved denitrification measurement method.

References

EPA. 1988. Boston Harbor Wastewater Conveyance System. Supplemental Environmental Impact Statement (SEIS). Boston: Environmental Protection Agency Region 1.

MWRA. 1991. Massachusetts Water Resources Authority effluent outfall monitoring plan: Phase I baseline studies. Boston: Massachusetts Water Resources Authority. Report ENQUAD ms-02. 95p.

MWRA. 1997a. Massachusetts Water Resources Authority effluent outfall monitoring plan: Phase II post discharge monitoring. Boston: Massachusetts Water Resources Authority. Report ENQUAD ms-44. 61p.

MWRA. 2003a. Briefing for the OMSAP workshop on ambient monitoring revisions: March 31-April 1, 2003. Boston: Massachusetts Water Resources Authority. Report ENQUAD ms-083. 91p.

MWRA. 2003b. Briefing for the OMSAP workshop on ambient monitoring revisions: June 18-19, 2003. Boston: Massachusetts Water Resources Authority. Report ENQUAD ms-085. 250 p.

NRC. 1990. Managing Troubled Waters: The Role of Marine Monitoring. National Research Council. National Academy Press, Washington, DC. 125pp.

2. BENTHIC NUTRIENT FLUX STUDY

Doug Hersh, Jane Tucker, and Anne Giblin

2.1 Introduction and Study Design

2.1.1 Objective and Scope

The Benthic Nutrient Flux Studies were initiated in 1990 to examine spatial and temporal trends of benthic processing of organic matter at selected stations in Boston Harbor and Massachusetts Bay.

The overall objectives of the studies have been to quantify sediment-water exchanges of oxygen, total carbon dioxide, and nutrients to define benthic-pelagic coupling in the harbor and bay. In addition, sediment indicators of organic matter loading and processing, such as organic carbon and pigment concentrations and redox conditions, have been studied.

Specific questions addressed by this study are:

⇒ How do the sediment oxygen demand, the flux of nutrients from the sediment to the water column, and denitrification influence the levels of oxygen and nitrogen in the water near the outfall?

⇒ Have the rates of these processes changed with relocation of the outfall?

Until late in 2000 the focus of these studies was on monitoring the recovery of the harbor as sewage treatment was improved, and in providing baseline information about all of these processes in Massachusetts Bay before the ocean outfall became operational. Delays in the construction of the outfall allowed for a much longer baseline period than originally planned (9 years instead of 3), during which we were better able to assess the natural variability in Massachusetts Bay.

In September 2000 Deer Island Treatment Plant effluent was diverted from Boston Harbor to the bay outfall. In 2001 monitoring of the harbor recovery continued and baseline monitoring of the bay ended. The emphasis changed to monitoring the response of the bay ecosystem to the relocation of the outfall.

2.1.2 Study design

Sediment cores are collected during four surveys in May, July, August, and October. This sampling strategy provides data across the approximate annual range of bottom water temperatures in both Boston Harbor and Massachusetts Bay, as well as providing information during the critical warmer months when the Bay water column is stratified. Sediment sampling stations (Figure 2-1) in Boston Harbor are stations BH02, BH03, BH08A, and QB01. Massachusetts Bay stations are stations MB01, MB02, MB03, and MB05.

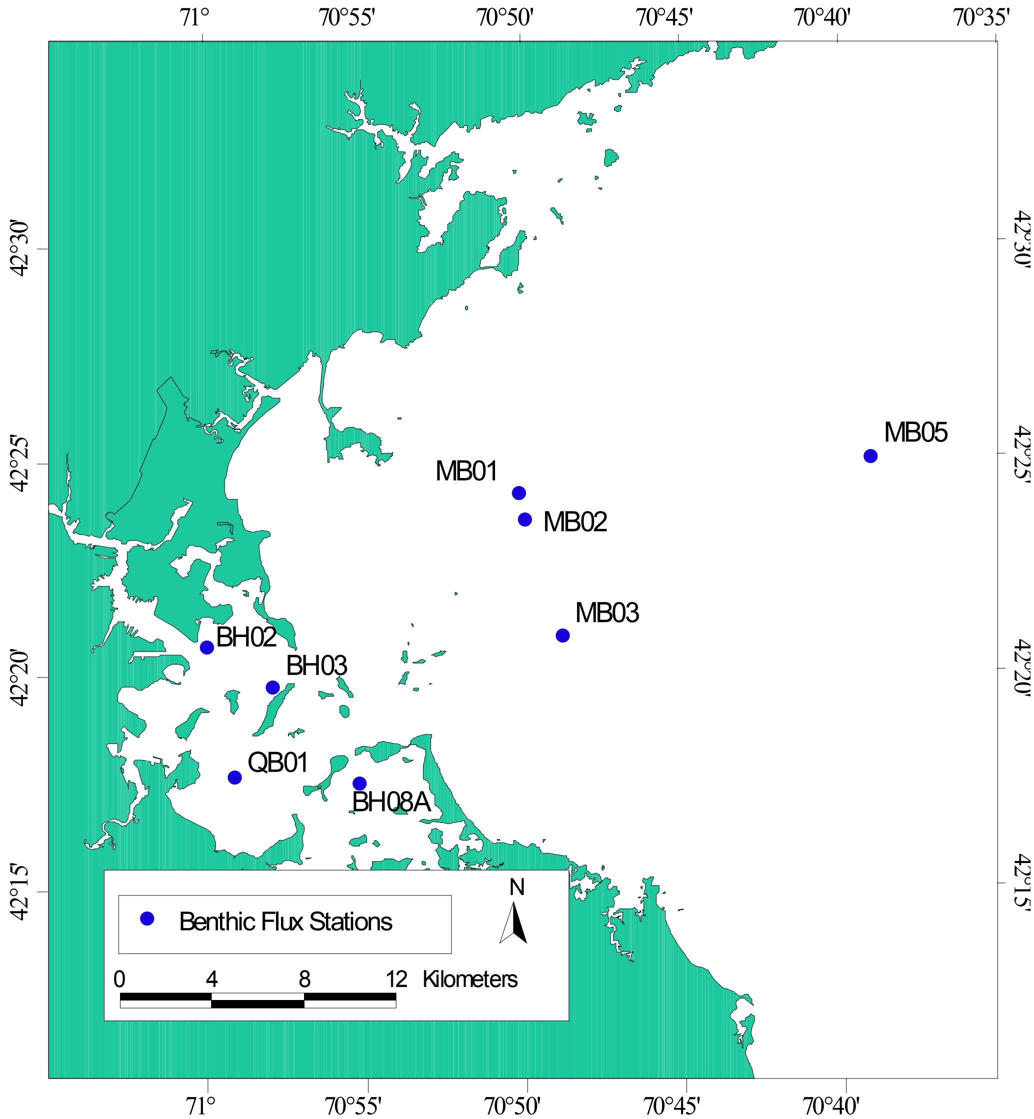


Figure 2-1 There are eight benthic flux stations, four in the bay and four in the harbor.

Sediment cores are returned to the laboratory, where flux incubations are performed on intact cores. Other cores are used for porewater analyses and measurements of solids (Table 2-1).

Denitrification fluxes are carried out on cores from harbor stations BH02 and BH03 during all four surveys, and from Stations MB02 and MB03 only during the May and October surveys, the time within our sampling season when bottom waters are typically the coolest and warmest, respectively, for the Bay. The May and October surveys may also capture the effects of the spring and fall phytoplankton blooms. We plan to use a new method to measure denitrification in upcoming surveys (See endnote, New Denitrification Method).[i]

Porewater profiles of nutrients, alkalinity, and dissolved sulfides are measured during only the July and August surveys, when bottom waters are warm and rates of benthic metabolism are high. Measurements made at this time allow the midsummer peak in free sulfide to be tracked if it appears during this period of peak activity. Samples are taken for grain size analyses in May and October.

All other parameters are measured at each station for each survey (Table 2-1). Detailed methods are in Tucker and Giblin 2002.

Table 2-1 Parameters measured by Benthic Nutrient Flux study.

Analysis	Sample Type (Number per Station)	Parameter	Method	Units	Reference	Frequency of Sampling
Flux	15 cm diameter core (2)	O_2	Electrode	μM	Hale, 1980	≥ 5 per flux
		Total CO_2	Coulometric CO_2 analyzer	μM	DOE, 1994	2(initial + final)
		NH_4	Spectrophotometric	μM	Solorzano, 1969	~ 5 per flux
		NO_2+NO_3	Flow injection analyzer	μM	Diamond, 1994	~ 5 per flux
		PO_4	Spectrophotometric	μM	Murphy and Riley, 1962	~ 5 per flux
		Si	Rapid Flow Analyzer	μM	Armstrong, 1951	~ 5 per flux
		Urea	Spectrophotometric or Rapid Flow Analyzer	μM	Price and Harrison, 1987	~ 5 per flux
	10.1 cm diameter core (1 oxic)	N_2	Gas chromatograph	μmoles m^{-2} d^{-1}	Kelly and Nowicki, 1993	~ 4 per flux
		O2	Gas chromatograph	μmoles m^{-2} d^{-1}	Kelly and Nowicki, 1993	~ 4 per flux
	10.1 cm diameter core (1 anoxic)	N_2	Gas chromatograph	μmoles m^{-2} d^{-1}	Kelly and Nowicki, 1993	~ 4 per flux
Porewater Nutrients	6.5 cm diameter core (1)	NH4	Spectrophotometric	μM	Solorzano, 1969	≥ 6 depth intervals
		NO2+NO3	Spectrophotometric	μM	Diamond, 1994	≥ 6 depth intervals
		PO4	Spectrophotometric	μM	Murphy and Riley, 1962	≥ 6 depth intervals
		Sulfide	Spectrophotometric	mM	Cline, 1969	≥ 6 depth intervals
		Si	Rapid Flow Analyzer	μM	Armstrong, 1951	≥ 6 depth intervals
		Urea	Spectrophotometric or Rapid Flow Analyzer	μM	Price and Harrison, 1987	≥ 6 depth intervals
		Alkalinity	Titration	Milli-equivalents	Edmond, 1970	≥ 6 depth intervals
Porewater Redox	6.5 cm diameter core (1)	pH	In-situ probe or electrode		Mitchell 1997 or Edmond 1970	≥ 6 depth intervals
		Eh	Probe	mV	Bohn, 1971	≥ 6 depth intervals
		Aparent RPD	Visual inspection	cm		One depth per core
Solids	6.5 cm diameter core (1)	Grain Size	Stacked sieves on Fritsch Analysette vibration table and popette/settling procedures	Percent dry weight	Folk 1974	Top 2-cm
	2.5 cm diameter core (1)	Porosity	Balance	g/mL	Giblin *et al*. 1994	1-cm intervals to 10 cm, 2-cm intervals thereafter
	2.5 cm diameter core (1)	Chlorophyll and Phaeophytin	Spectrophotometric	μg/mL	Lorenzen, 1967	1-cm intervals to 5 cm
	2.5 cm diameter core (1)	Total Organic Carbon and Total Nitrogen	Elemental analyzer	Percent dry weight	Kristensen and Andersen, 1987	Top 2 cm
Seawater	*In situ*	O_2, Salinity, Temperature	Hydrolab multiparameter system	mg/L, PSU, C	Hale, 1980	1 set of measurements at each station

2.2 Monitoring Results

2.2.1 Bay Summary

Detailed results are in Tucker *et al*., 2003. Most sediment fluxes in 2002 and 2001 were no higher than levels observed during the baseline period and for oxygen, somewhat lower than observed in 1999. Oxygen (O_2) fluxes, or sediment oxygen demand, at MB01 were higher in 2002 than in 2001.

Dissolved inorganic nitrogen fluxes are more variable than sediment oxygen demand but 2002 fluxes were within baseline values except for ammonium uptake, which was higher at two of the four Massachusetts Bay stations. Silicate fluxes were quite low, perhaps reflecting the failure of the spring bloom in 2001 and 2002.

Sediment redox showed no changes in 2002 from the baseline period.

At two of the four Massachusetts Bay stations sediment total organic carbon values were higher than in recent years. While the values were still within baseline values, the downward trend we had observed at these two stations has reversed (Tucker *et al*., 2003).

In 2002, there was no indication of increased sediment oxygen demand or increased nutrient fluxes from nearfield sediments. In fact, fluxes were generally in the low end of baseline observations or were lower.

Although the rates of sediment oxygen demand would be high enough to affect the seasonal drawdown of oxygen in the water column in the nearfield if the water column were stagnant, the rate of water renewal from the Gulf of Maine is sufficient to nearly completely override the effect of local benthic metabolism. The renewal rate of the bottom water, which is determined by wind and other climatological factors, determine the timing and strength of the seasonal oxygen drawdown (Geyer *et al*. 2002).

2.2.2 Harbor summary

Northern Harbor – Respiration rates at station BH03 in 2001 and 2002 continued a downward trend; the rates in 2002 were the lowest yet measured. Rates at BH02 were also the lowest ever measured. This station shows more variation but overall the trend is still downward.

Southern Harbor – Respiration rates at Quincy Bay station QB01 remain unchanged from baseline years. Rates decreased somewhat at BH08A in 2001 but there was a very large drop in 2002.

It will take another 1-2 years to say with certainty that there is a harbor response to outfall relocation but the data are suggestive.

2.3 Proposed change: Eliminate urea fluxes

Urea flux has not proved useful as a measurement of nitrogen flux components. Of the three nitrogen nutrient fluxes we routinely measure, combined nitrate and nitrite (NO_3+NO_2), ammonium (NH_4), and urea, NH_4 accounts for about 2/3 in Boston Harbor and almost all of the nitrogen nutrient flux in Massachusetts Bay and Stellwagen Basin (Figure 2-2). NO_3+NO_2 flux typically accounts for about 1/3 of the net DIN+urea flux in Boston Harbor and is only a minor component of the net flux in Massachusetts Bay and Stellwagen Basin. Half of the urea flux values are between –5 and 5 percent of total nitrogen nutrient flux.

Both NH_4 and NO_3+NO_2 show decreased ranges of flux values from Boston Harbor to Massachusetts Bay to Stellwagen Basin (Figure 2-2) in response to what are presumably environmental gradients. Urea flux shows no such changes from in-shore to offshore locations. This appears to be true both before and after outfall relocation (Figure 2-2).

These fluxes are calculated from samples taken at five (maximum) time points during each sediment core incubation. Whereas NH_4 and NO_3+NO_2 fluxes typically reflect clear linear changes in concentration of these nutrients over the time course of the incubation, this is rarely the case for urea (Table 2-2). Urea concentrations measured over the same time course are typically low and variable, resulting in poor r-squared values for the calculated flux.

Table 2-2 R-squared values of nitrogen fluxes measured since 1998

Nutrient flux	Average of R-square values	Standard deviation of R-square values
NH_4	0.86	0.23
NO_3+NO_2	0.81	0.27
Urea	0.40	0.34

We propose the elimination of urea flux measurements because these fluxes have never been a significant portion of net nitrogenous nutrient fluxes, and they have not added to our understanding of benthic responses to changes in nutrient and organic loading in the Boston Harbor/Massachusetts Bay system.

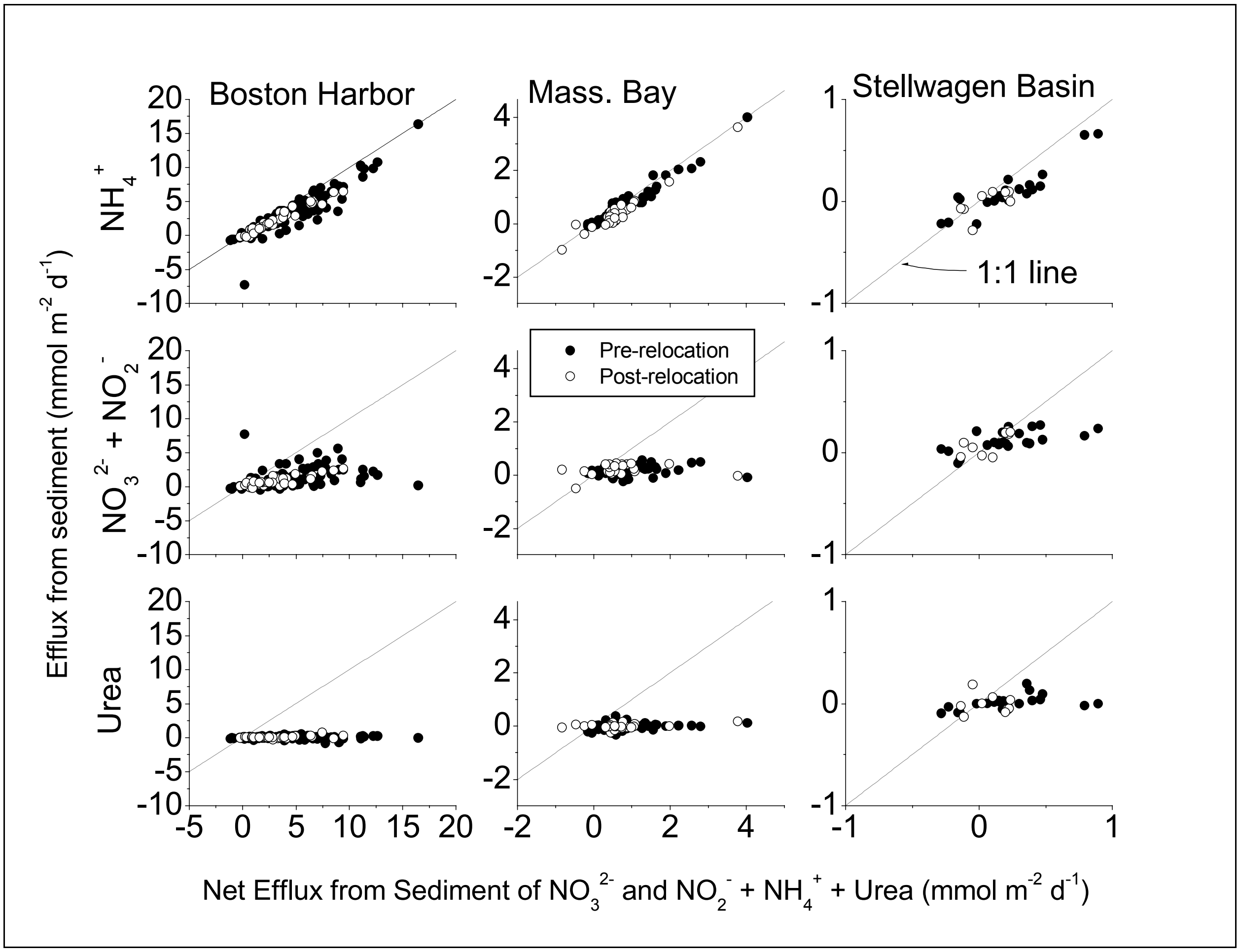


Figure 2-2 Fluxes of urea, nitrate+nitrite, and ammonium compared to their net flux. Each point represents the average flux for a station on a given day. Data shown are from 1994 to 2002 (excluding 1995, which had atypical nutrient flux ratios). Note scale changes on both axes.

2.4 Proposed change: Eliminate porewater profiles

Currently, porewater profiles of nutrients, alkalinity, and dissolved sulfides (H_2S) are measured during July and August surveys only, when bottom waters are warm and rates of benthic metabolism are high. In the past, we have not used porewater profiles as a routine monitoring tool, but rather to help understand the patterns we see in fluxes, which we measure directly by incubating sediment cores under controlled conditions in the laboratory.

In Boston Harbor, where there is significant spatial and temporal variation in flux rates and metabolism, porewater profiles have helped to elucidate mechanisms underlying variation of fluxes. For example, porewater profiles can help distinguish between two conditions under which sediment oxygen demand (SOD) and NH4 fluxes are high. High fluxes may be related to organic matter inputs, which would be reflected in low redox values and high porewater concentrations of H2S and NH4 (see station BH02 in Table 2-3 and Figure 2-3). In contrast, they may be related to high rates of bioturbation, which would result in more oxidized sediments and lower porewater NH4 and H2S (see station BH03A in the same table and figure). In the first case, the condition might be viewed as a "worsening" of the system, whereas the second case might be viewed as an improvement.

While porewater profiles can be helpful, other monitoring measurements provide much of the same information. In the case shown in Table 2-3 and Figure 2-3, redox profiles, the measurement of nutrient fluxes on incubated cores, and the observation of dense mats of tube-dwelling *Ampelisca* amphipods in the cores from station BH03A are sufficient to understand the differences between the two cores. Low redox values in the core without amphipods correspond to high porewater sulfides. The differences between the cores in NH_4 and PO_4 porewater concentrations are mirrored by differences in flux values. These measurements are all consistent with the effects of bio-irrigation by amphipods. Measurements of porewater nutrients are not required either to quantify the flux rates or identify sediments with low redox status because the other measurements we make provide redundant information.

In Massachusetts Bay, where we have observed no changes due to the outfall relocation in nutrient flux rates and metabolism, porewater profiles are even less important. In contrast to Boston Harbor, where low Eh and high sediment H_2S values of 1 to 5 mM are common, Massachusetts Bay and Stellwagen Basin sediment redox values tend to be high and H_2S concentrations correspondingly low (<0.1 mM) deep into the sediment (Figure 2-4). Thus, for Massachusetts Bay, porewater profiles have not yet yielded information necessary for the interpretation of the low flux rates and high Eh values we routinely observe.

We suggest elimination of porewater measurements in both Boston Harbor and Massachusetts Bay unless significant changes in fluxes are observed or very negative (<-150 mV) Eh values are recorded in the future. We would continue to record redox and pH profiles, which are measured with probes on intact cores. If such low redox values or increases in fluxes were observed at a particular location in Massachusetts Bay, we would evaluate whether to do porewater profiles at that station on a subsequent survey to identify the mechanism involved in the change. Porewater measurements, which are time intensive, are appropriate as an ancillary technique to help us understand changes we see using other methods.

Table 2-3 Comparison of two Boston Harbor stations during periods of high sediment oxygen demand.

Date	Station	Temp. (C)	O_2 flux*	NH_4 flux*	NO_3 flux*	Si flux*	PO_4 flux*	*Ampelisca* mat
Aug. 1997	BH03A	18.3	-168	3.5	5.6	12.8	0.35	present
May 1993	BH02	12.3	-107	12.2	-0.37	10.3	7.8	absent

* all flux values in millimoles m^{-2} d^{-1}

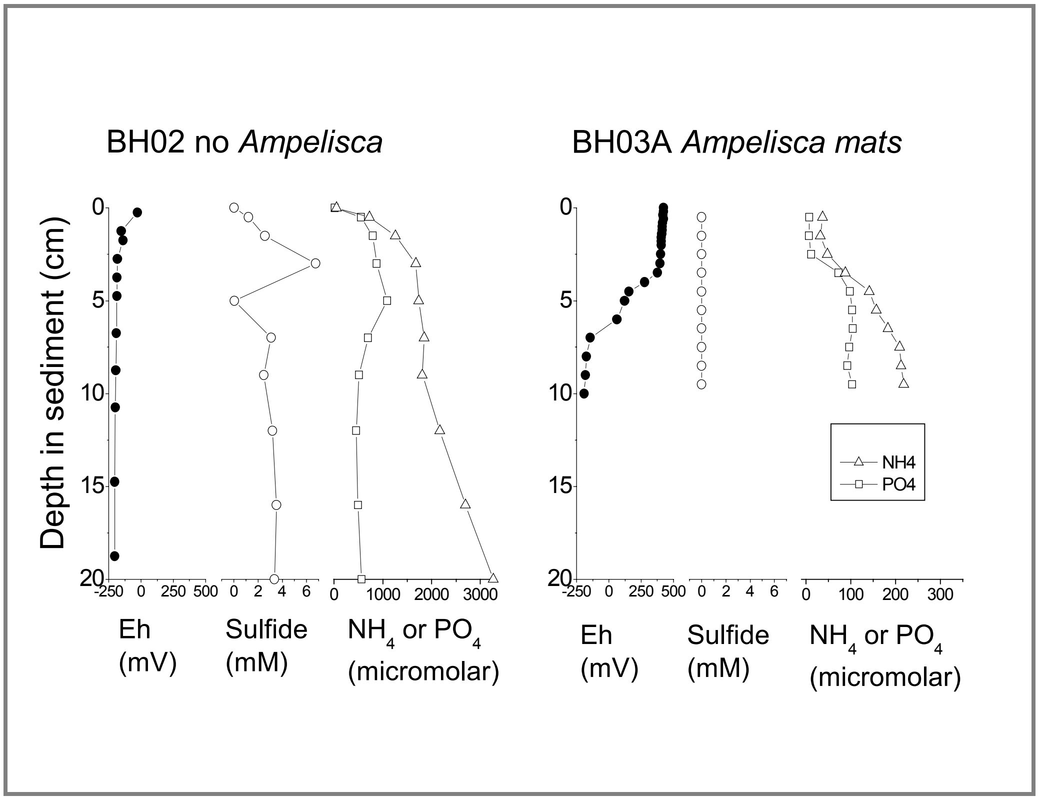


Figure 2-3 Porewater profiles of sediment redox, dissolved sulfide, and ammonium at two selected Boston Harbor stations with high sediment oxygen demand and ammonium flux. Refer to Table 2-3 for flux rates and ambient conditions at the two stations shown. Note change in scale for NH_4 and PO_4.

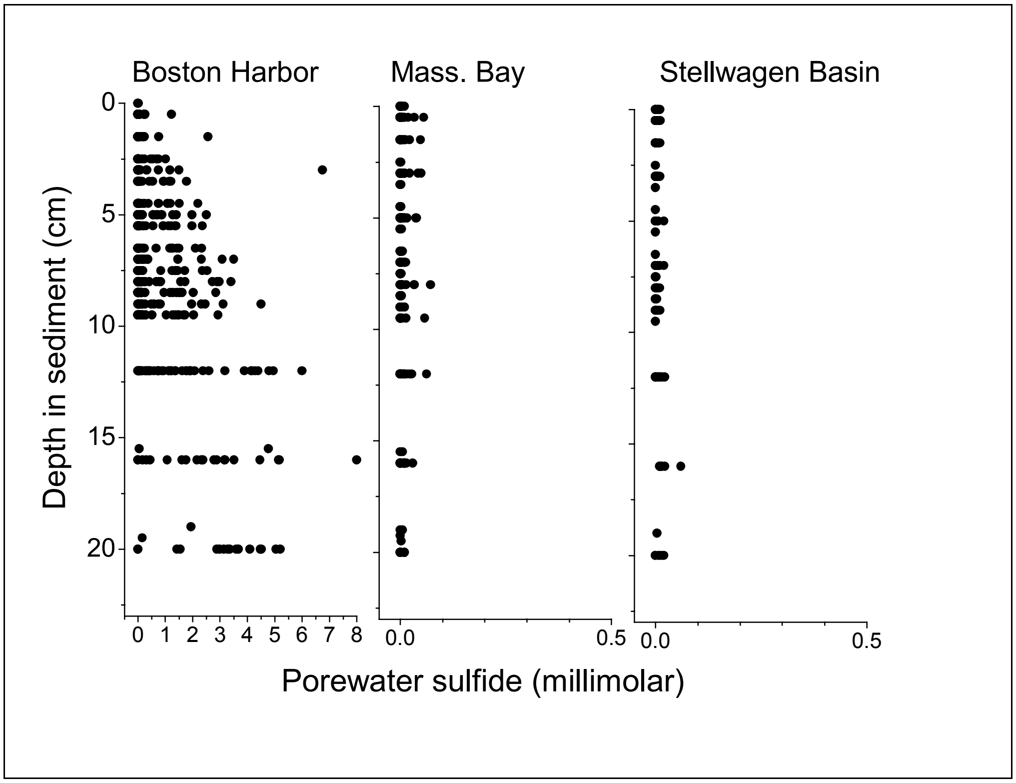


Figure 2-4 Concentration of sulfides in sediment of all Boston Harbor stations, Massachusetts Bay stations (MB01, MB02, and MB03), and Stellwagen Basin station MB05. Data are from 1993 to 2002. Note change in x-axis scale among plots.

2.5 Future changes under consideration

In 2005 MWRA will again review the flux results. If findings are comparable to previous years we will evaluate whether this study should be continued. If so, we will determine, in consultation with OMSAP, which measurements are appropriate to address remaining questions about the impact of outfall relocation on nutrient flux benthic metabolism in Massachusetts Bay.

References

Geyer WR, Libby PS, Giblin A. 2002. Influence of physical controls on dissolved oxygen variation at the outfall site. Boston: Massachusetts Water Resources Authority. Letter Report ENQUAD. 20 p.

Tucker J and Giblin A. 2002. Combined Work/Quality Assurance Plan (CWQAPP) for Benthic Nutrient Flux Studies: 2002-2005. Boston: Massachusetts Water Resources Authority. Report ENQUAD ms-077. 46 p.

Tucker J., Kelsey S., Giblin A., and Hopkinson C. 2003. 2002 Annual benthic nutrient flux monitoring report. Boston: Massachusetts Water Resource Authority. Report ENQUAD 2002-xx. 52 p

[i] **New denitrification method**

Since 1992, we have used a method to measure denitrification developed by Barbara Nowicki at University of Rhode Island. A detailed description of sampling and measurement methods for this technique is given in Nowicki *et al*. (1997) and in the CW/QAPP (Tucker and Giblin, 1998). Briefly, for each estimate of N_2 flux, two sediment cores are incubated at ambient temperature. Prior to the flux incubation, the overlying water and headspace in one core (the oxic core) is sparged with a mixture of 80% Helium (He):20% O_2 in order to maintain an oxic environment while lowering the background level of N_2 within the core. The overlying water and headspace in the other core (the anoxic core) is sparged with 100% He to remove both O_2 and N_2. The anoxic core provides an abiotic control for N_2 diffusing from the porewater and conditions under which coupled nitrification/denitrification is prevented. During the incubation, N_2 gas within the headspace of both cores is monitored by drawing off samples of the headspace and analyzing it on a gas chromatograph with a thermal conductivity detector. Denitrification is calculated as the difference in N_2 production in the oxic and anoxic cores.

A newer technique for measuring denitrification uses a quadrupole mass spectrometer equipped with a membrane inlet (membrane inlet mass spectrometer or MIMS) to precisely measure N_2/Argon (Ar) ratios of dissolved gases in water samples (Kana *et al*., 1998). Dinitrogen gas concentrations are affected by both biological and physical processes, whereas Ar is affected only by physical processes. Deviations from equilibrium ratios of these two gases therefore reflect biological processes acting on the N_2. The mass spectrometer is capable of measuring very small deviations in this ratio, thereby providing a very sensitive and precise method for measuring denitrification. Whereas gas chromatography offers a precision of measurement for gas concentrations on the order of 0.3-1%, the mass spectrometer yields a precision of 0.05% for gas ratios.

Importantly for our benthic flux studies in Boston Harbor and Massachusetts Bay, samples for dissolved gas analysis (DGA) are taken from the same cores as are used for flux measurements, allowing for direct comparison of fluxes from a given core. Four to five samples are taken over the incubation time course, simultaneously with the nutrient flux samples.

The MBL laboratory has run a limited number of samples comparing the GC and MIMS technique (Figure i-1). Results indicate that the two techniques are not directly comparable. Most of the GC results are higher than MIMS. Other GC results are lower, perhaps reflecting variation in one or both techniques.

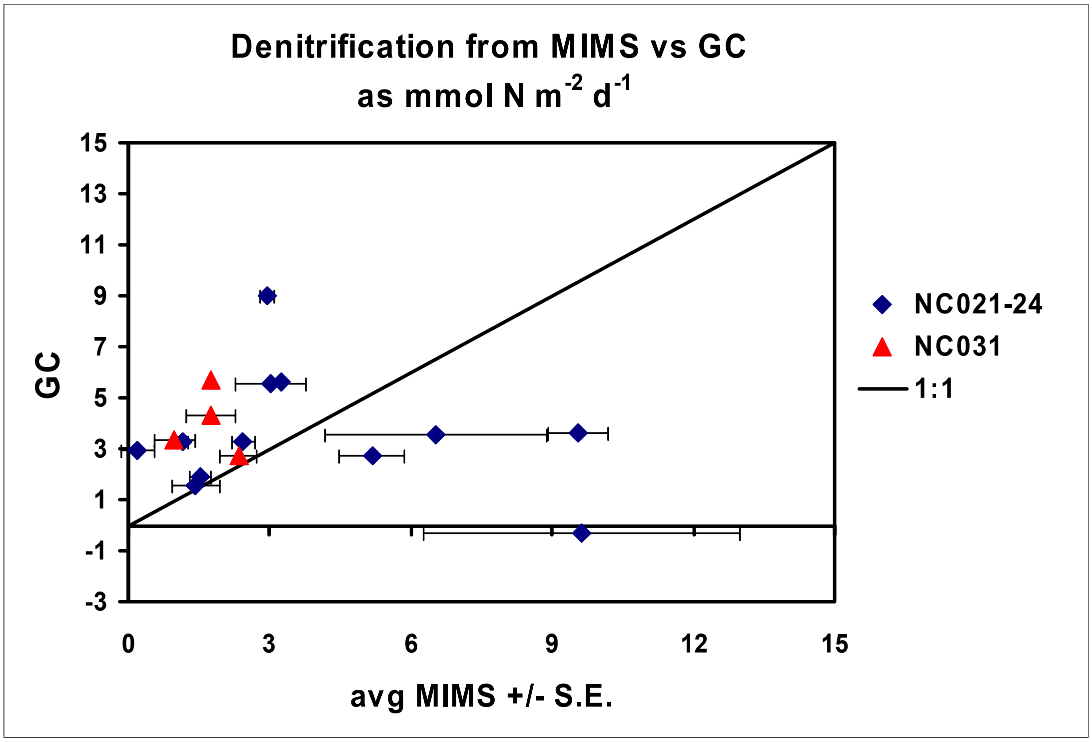


Figure i-1 Comparison of denitrification results using the MIMS technique versus the GC technique. MIMS data are the average of two measurements and the error bars represent standard error.

For a number of reasons, the GC technique may be expected to have higher or more variable results than the MIMS technique. One reason is that bio-irrigation by benthic invertebrates leads to faster abiotic degassing of N_2 from sediments in the oxic core than in the anoxic core (in which the animals would have been killed). Another reason is that during the long incubations used for the GC technique (typically 5 days), NH_4 builds up in both the porewater and the overlying water. Higher porewater ammonium leads to higher nitrification, which in turn leads to higher coupled nitrification/denitrification (Giblin, unpublished data). The MIMS method has a shorter incubation time of 2-3 days, which should result in less build-up of NH_4 over the course of the incubation.

The GC technique is also very labor intensive. It takes about 7 person hours for each denitrification rate results. Time constraints limit the number of measurements, which limits spatial and temporal coverage and our ability to do replicate measurements. The MIMS technique, in comparison, takes under one fourth of the time for each denitrification result.

The MIMS technique does require greater care in sample handling and has stricter tolerances for changes in temperature and pressure. Nevertheless, its benefits out-weigh these difficulties enough for this technique to have become the preferred technique.

MWRA plans to change to the newer MIMS technique, which allow us more flexibility in our sampling design and allow us to do more measurements if they are required.

References

Kana, T.M., M.B. Sullivan, J.C. Cornwell and K. Groszkowski. 1998. Denitrification in estuarine sediments determined by membrane inlet mass spectrometry. *Limnology and Oceanography* 43: 334-339.

Nowicki, B.L, J.R. Kelly, E. Requintina, D. Van Keuren. 1997. Nitrogen Losses Through Sediment Denitrification in Boston Harbor and Massachusetts Bay. Estuaries. 20:626-639.

Tucker J and Giblin A. 2002. Combined Work/Quality Assurance Plan (CWQAPP) for Benthic Nutrient Flux Studies: 2002-2005. Boston: Massachusetts Water Resources Authority. Report ENQUAD ms-077. 46 p.

Tucker J, Kelsey, S, Giblin, A, and Hopkinson C. 2003. 2002 Annual Benthic Flux Monitoring Report. Boston: Massachusetts Water Resources Authority. Report ENQUAD 2003-08. 52 p.

3. SEDIMENT INFAUNAL MONITORING

Kenneth E. Keay, Eugene D. Gallagher, James A. Blake, Nancy J. Maciolek

3.1 Background

3.1.1 Sedimentary environment

The sea floor of Massachusetts and Cape Cod bays was originally shaped by the glaciers, which sculpted the bottom and deposited debris, forming knolls, banks, drumlins, buried river channels and other features. Within Massachusetts Bay, the texture of the sea floor ranges from mud in depositional basins to coarse sand, gravel, and bedrock on topographic highs. The area around the outfall is marked by underwater drumlins, which are elongated hills about 10 meters high, with crests covered by gravel and boulders. Between these drumlins are small depositional depressions. The major regional long-term sinks for fine-grained sediments include Boston Harbor, Cape Cod Bay, and Stellwagen Basin (USGS 1997a, 1998).

Most of the sediment mass transport in the region occurs primarily during storms. Typically, waves during storms with winds from the northeast resuspend sediments, which are transported by shallow currents from western Massachusetts Bay toward Cape Cod Bay and by deeper currents to Stellwagen Basin. Cape Cod Bay is partially sheltered from large waves by the arm of Cape Cod, and storm waves are rarely large enough to resuspend sediments in Stellwagen Basin, which is the deepest feature in the region. Therefore, sediments deposited in these locations are likely to remain.

3.1.2 Environmental concerns

Within Boston Harbor, studies of the sediments have documented recovery following the cessation of sludge discharge, improvements to Combined sewer overflow systems, and improved sewage effluent treatment (Bothner *et al*. 1998, Lefkovitz *et al*. 1998, Kropp *et al*. 2002). However, relocating the outfall raised concerns about potential effects of the relocated discharge on the fauna inhabiting the offshore sea floor. These concerns focused on three issues: eutrophication and related low levels of dissolved oxygen, accumulation of toxic contaminants in depositional areas, and smothering of animals by particulate matter. MWRA's sediment monitoring addresses all three concerns. This briefing package (and the discussions at the July 24, 2003 OMSAP workshop) focuses on infaunal community monitoring. Sediment contaminant monitoring and the hard-bottom special study were discussed in detail at the April 1, 2003 OMSAP workshop and are summarized in that briefing package, as are the effluent monitoring results (MWRA 2003a). The briefing package for the June 2003 OMSAP workshops contained additional detail on effluent nutrient loads and speciation relative to eutrophication concerns (MWRA 2003b).

If transfer of the nutrient loads to offshore were to cause eutrophication, depressed levels of dissolved oxygen (DO) or a change in the input of food quality or quantity to the benthos might affect bottom communities. Increasing the amount of particles and organic matter delivered to the bottom could disrupt normal benthic community structure in the vicinity of the discharge, even in the absence of depression of water column DO.

Although the Deer Island Treatment Plant was designed to keep effluent solids and biochemical oxygen demand low, EPA's 1988 Environmental Impact Statement (EPA, 1988) predicted increased organic carbon deposition to nearby sediments resulting in faunal changes. One factor in these predicted increases was the expectation that the outfall would be discharging primary-treated effluent for five years, but modest increases were predicted even for secondary effluent. However, construction delays meant that the outfall did not come on-line until after the effluent was receiving secondary treatment, which effectively removes solids and BOD. The effluent monitoring data shown in Section II of the April briefing package (MWRA 2003a) confirm that the DITP discharges very low concentrations of BOD and TSS, compared both to concentrations projected in the SEIS and to the primary-treated effluent previously discharged to Boston Harbor.

Thus, as was discussed for effluent contaminant loadings at the sediment contaminant workshops in April, the relatively intense sediment monitoring in the Ambient Monitoring Plan (MWRA 1991, 1997) was designed to measure impacts from organic loadings that are lower than were projected.

When EPA prepared the SEIS, it was not possible to model the outfall's effects on primary production. If production was excessive, this could ultimately cause bottomwater DO depressions or increased organic matter deposition to the sediments. Later, as the Bays Eutrophication Model was developed, model runs projected that during summer stratified conditions the maximum deposition of organic matter is somewhat west of the outfall, in the same area as but of lower magnitude than deposition when a primary discharge in Boston Harbor is modeled (Hydroqual and Normandeau, 1995)(Figure 3-1)

As the monitoring results presented here will show, two years of post-discharge sampling have found no evidence of acute outfall-related impacts on sediment infaunal communities, either in the nearfield or in the farfield. MWRA believes that remaining concerns can be addressed with a more cost-effective program, beginning with the 2004 field sampling. Under our proposal, MWRA would continue to obtain infaunal data at all stations currently sampled every year. Two nearfield stations will continue to be sampled annually; remaining stations will be randomly split into two groups sampled in alternate years.

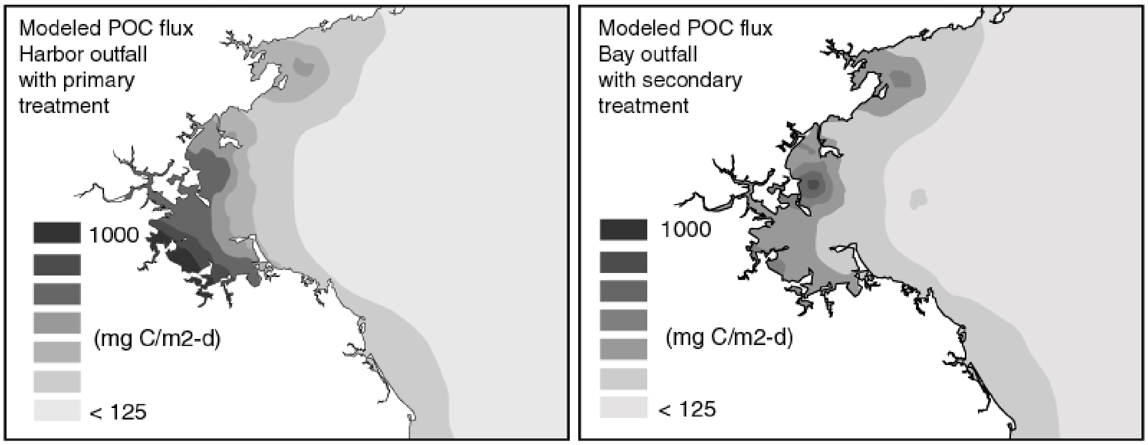


Figure 3-1 Modeled August particulate organic carbon deposition for (left) primary effluent discharged into Boston Harbor and (right) secondary effluent discharged at the Bay outfall. After Hydroqual and Normandeau, 1995.

3.2 Monitoring Design

Sediment monitoring was designed to address the following questions:

Sediment contamination and tracers

→ What is the level of sewage contamination and its spatial distribution in Massachusetts and Cape Cod Bays sediments before discharge through the new outfall?[1]

→ Has the level of sewage contamination or its spatial distribution in Massachusetts and Cape Cod Bays sediments changed after discharge through the new outfall?[1]

→ Have the concentrations of contaminants in sediments changed?[1]

Benthic communities

→ Has the soft-bottom community changed?

→ Have the sediments become more anoxic; that is, has the thickness of the sediment oxic layer decreased?

→ Are any benthic community changes correlated with changes in levels of toxic contaminants (or sewage tracers) in sediments?

→ Has the hard-bottom community changed?[1]

3.2.1 Sediment sampling

The infaunal monitoring consists of annual, mid-August sampling of sediments at 31 stations throughout Massachusetts and Cape Cod Bays (Figure 3-2, Figure 3-3). The nearfield sampling design evolved substantially in its early years as MWRA and the Outfall Monitoring Task Force evaluated how best to monitor infaunal communities in the area given the substantial between-station spatial heterogeneity in grain size and in infaunal communities identified in 1992 sampling (Blake *et al*. 1993), and then in light of the appreciable temporal heterogeneity in grain size and communities at stations observed between 1992 and 1993, which resulted from sediment

[1] Discussed during April 1 2003 OMSAP workshop

transport caused by the December 1992 “no-name” storm (Coats *et al.* 1995). The resulting design (in place since August 1994) is shown in Table 3-1. Infaunal samples are collected with a 0.044 m^2 Young-Van Veen benthic grab, sieved on 300*u*m mesh in the field, fixed in formalin, then transferred to alcohol and stained with Rose Bengal in the lab. Animals are sorted out of the samples, identified and counted. Estimates of the apparent redox potential discontinuity depth (RPD) are made on all grabs. Western Massachusetts Bay stations FF10, FF12, and FF13 are treated as nearfield stations in all analyses and evaluations.

Table 3-1 Station types and measurements for infaunal monitoring

Station type	Infauna	Grain Size[1]	Contaminants[2]	SPI	Stations	n
Replicated nearfield	3	2	2	Yes	FF10, FF12, FF13, NF12, NF17, NF24	6
Unreplicated nearfield	1	1	1	Yes	NF02, NF04, NF05, NF07, NF08, NF09, NF10, NF13, NF14, NF15, NF16, NF18, NF19, NF20, NF21, NF22, NF23	17
Farfield	3	2	2	No	FF01A, FF04, FF05, FF06, FF07, FF09, FF11, FF14	8
1 Includes measurements of total organic carbon and *Clostridium perfringens*						
2 As previously proposed, in 2003 and 2004 contaminant measurements will be restricted to stations NF12 and NF17						

3.2.2 Sediment profile imaging (SPI)

Multiple sediment profile images are collected at the 23 western Mass Bay stations (Table 3-1, Figure 3-2). These images are analyzed to determine evidence for recent sediment transport, estimate the grain size and depth of the RPD, presence of infaunal tubes and feeding structures, as well as a number of other parameters. The rapid turnaround (preliminary RPD estimates are available within one week of the survey) of the SPI study provides a useful supplement to the detailed data obtained from infaunal samples, whose analysis requires months.
Details of MWRA’s benthic monitoring can be found in the project work plan (Williams *et al*. 2002). Results of MWRA’s benthic monitoring have been presented in a series of annual interpretive reports, e.g. Blake *et al*. (1998), Kropp *et al*. (2001, 2002).

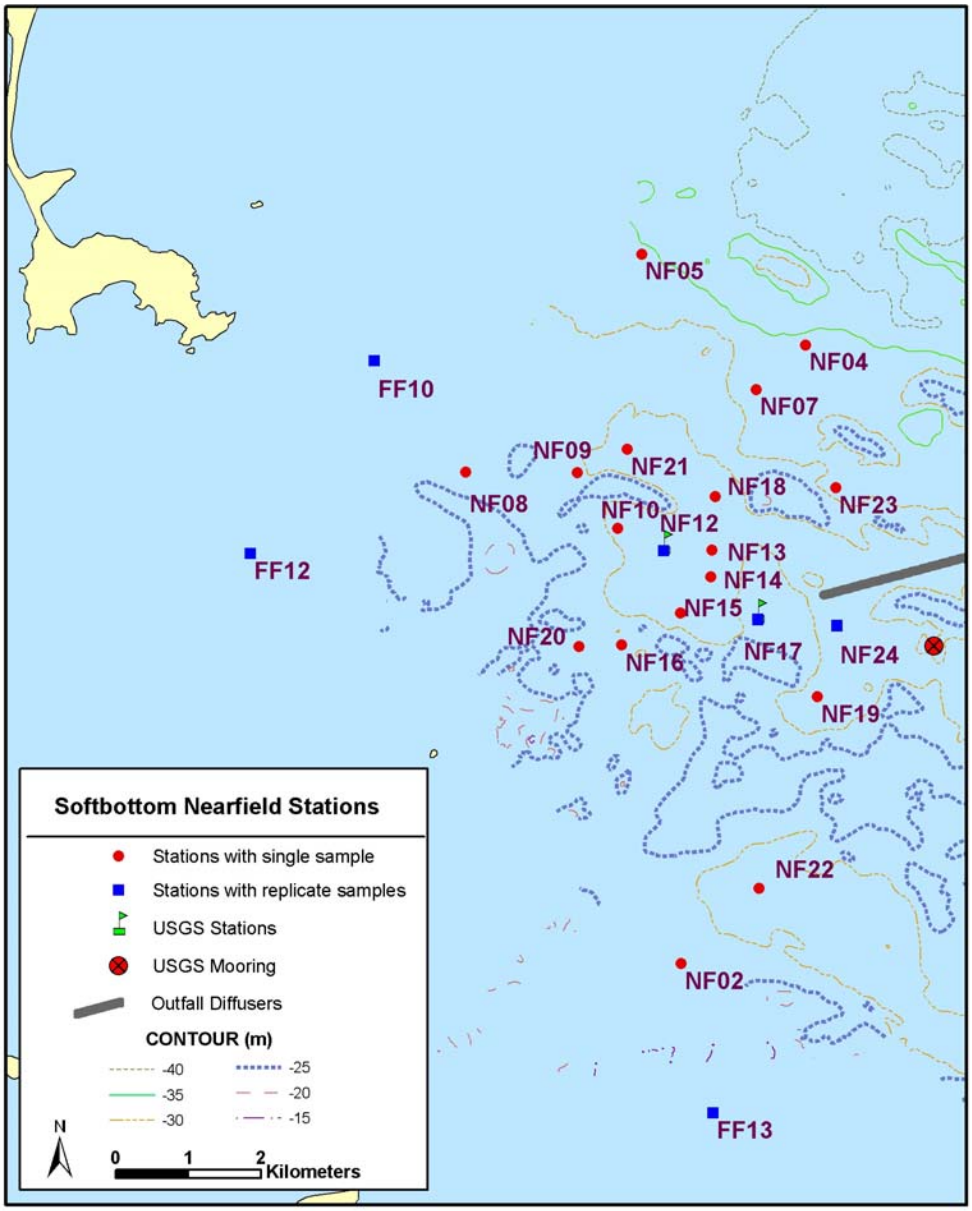


Figure 3-2 Locations of nearfield soft-bottom stations, including stations sampled by USGS.

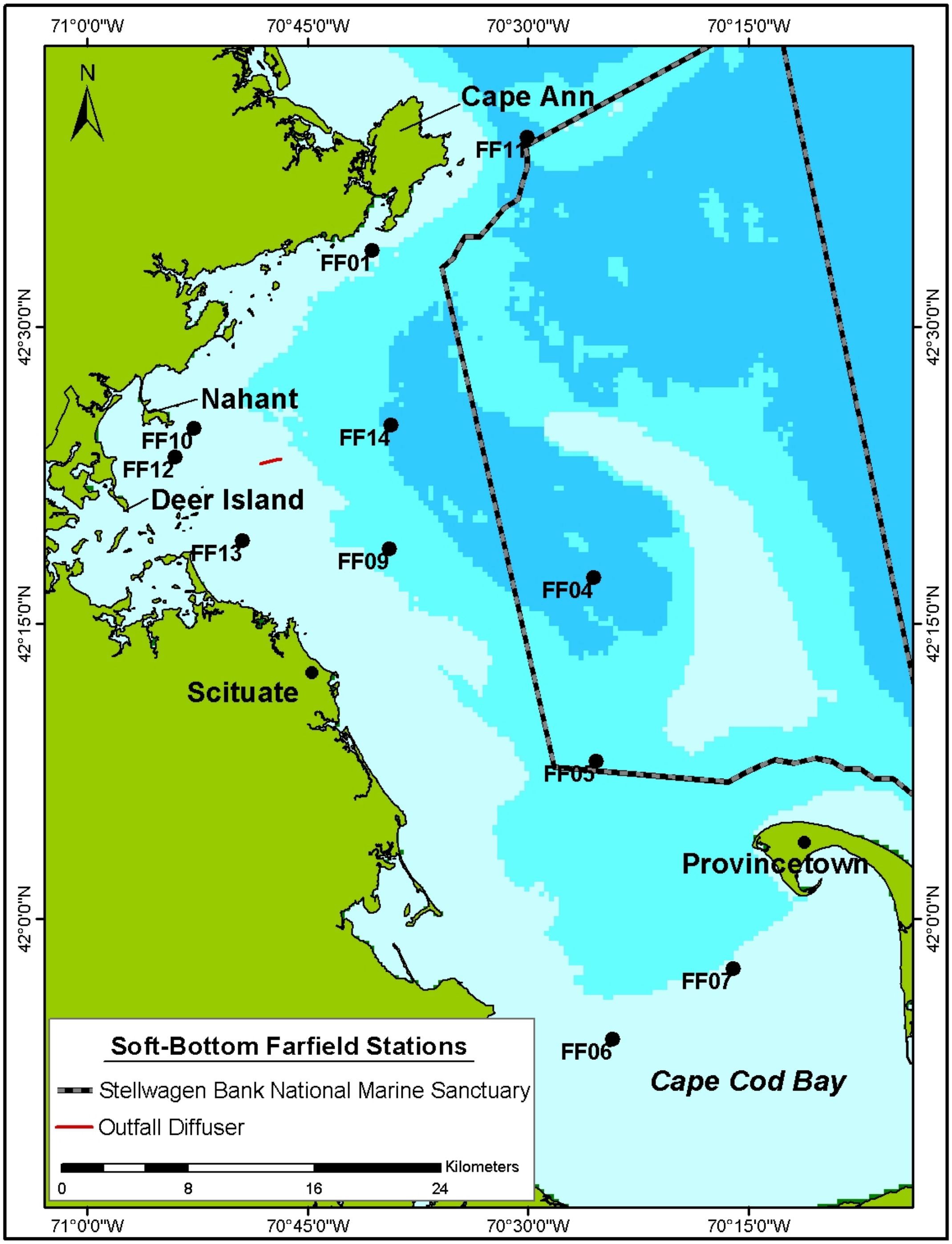


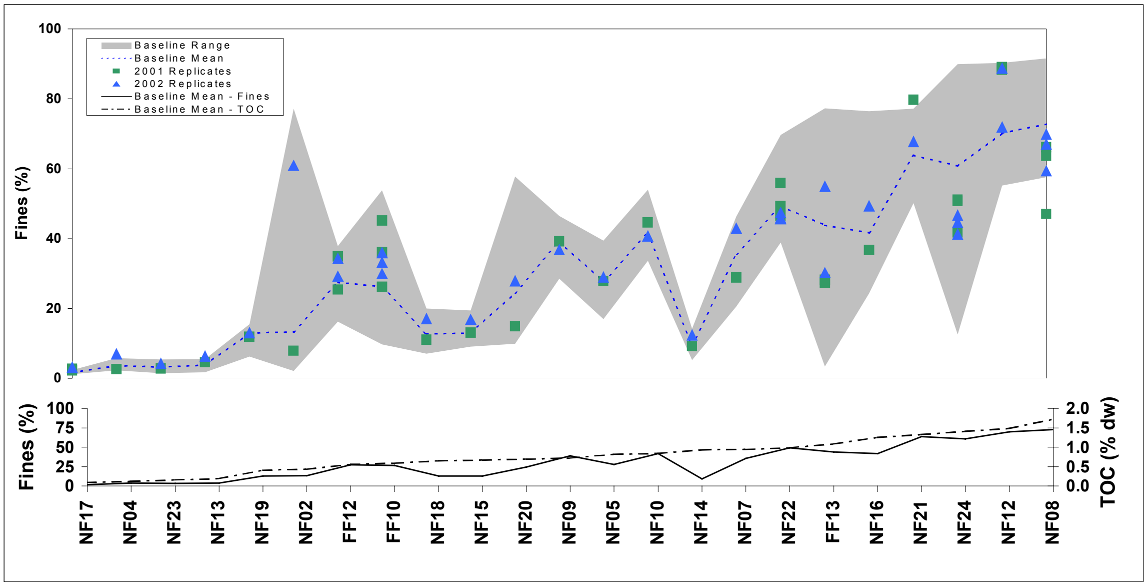
Figure 3-3 Locations of farfield soft-bottom stations

3.3 Monitoring Results

This section summarizes recent results which typify the findings of the monitoring. Western Massachusetts Bay farfield stations FF10, FF12, and FF13, which are within 7 km of the outfall, are included with the 20 NF stations in all nearfield analyses (including threshold computations). When "nearfield" and "farfield" comparisons are presented, results from those 23 stations are contrasted with those from the 8 remaining farfield stations. All replicated stations (8 farfield stations plus replicated nearfield stations FF10, FF12, FF13, NF12, NF17, and NF24) are included in some "regional" data plots.

3.3.1 Sedimentary environment

Baseline sampling at nearfield stations documented that the depositional environments in western Massachusetts Bay are quite heterogenous, with the baseline means for individual stations ranging from less than 5% silt+clay to over 70% (Figure 3-4). The farfield stations tend to be more consistently fine-grained than is the nearfield, especially the deepwater stations FF04, FF05, FF11, and FF14 (Figure 3-5). Organic carbon concentrations track grain size closely in both the nearfield and farfield, with correlation coefficients for baseline data (1992-2000) of 0.82 for the nearfield and 0.93 for the farfield (Kropp *et al*. 2002).



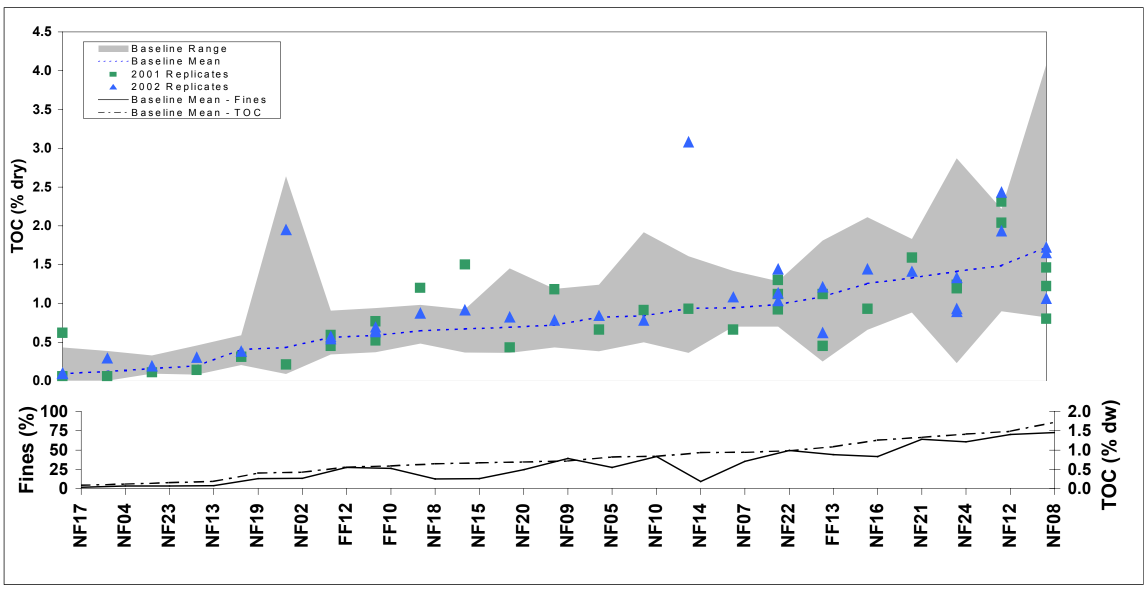
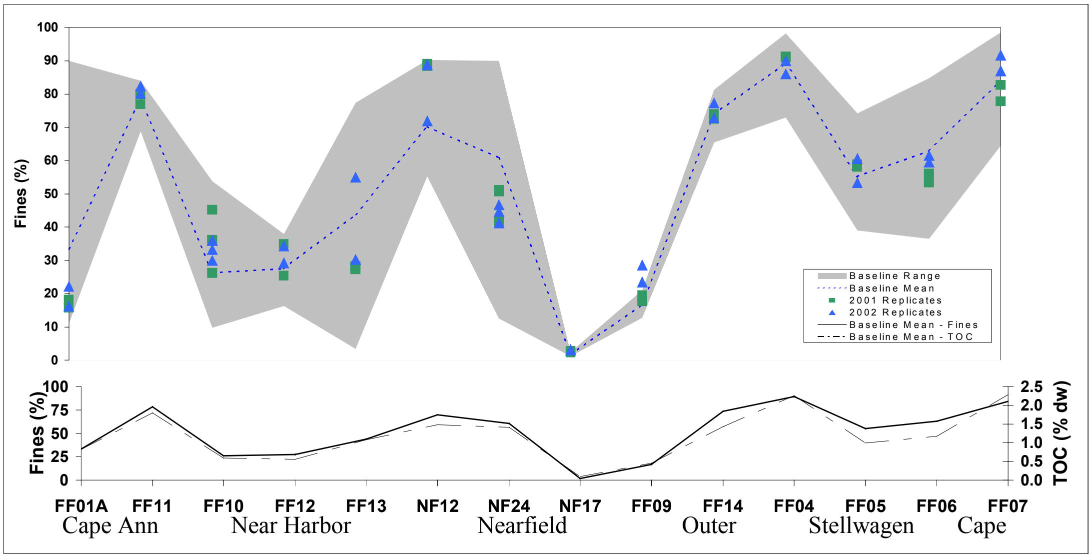


Figure 3-4 Percent fines (top) and TOC (bottom) for each nearfield station sampled in 2001 (squares), 2002 (triangles) and the range of values occurring during the baseline period (gray band). The baseline mean value is indicated (dashed line within gray band).



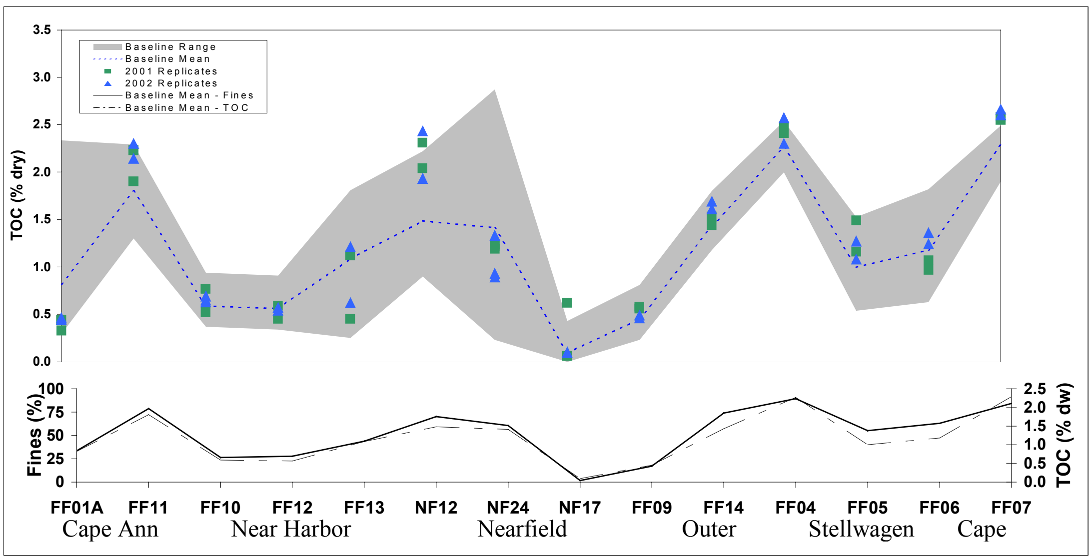


Figure 3-5 Percent fines (top) and TOC (bottom) for each regional station sampled in 2001 (squares), 2002 (triangles) and the range of values occurring during the baseline period (gray band). The baseline mean value is indicated (dashed line within gray band). Stations are presented in order of location relative to the outfall, from north to south. Baseline station mean values for TOC and percent fines, plotted by station, are shown in the sub-plot.

Storm-driven transport of fine sediments is a major factor impacting the shallower depositional environments in Massachusetts Bay (Knebel and Circe 1995) and the infaunal communities they support. The largest such event yet observed followed the major northeast storm of December 11-16, 1992 (Bothner *et al*. 2002). Sediment concentrations of clay in the upper 0.5 cm peaked at station NF12 in USGS samples collected after that storm (Figure 3-6) which caused sustained wave heights in excess of 7m in the vicinity. Suspended sediment and turbidity measurements from the USGS hydrographic mooring also document the importance of storms to sediment transport in the nearfield (Bothner *et al*. 2002, Butman *et al*. 2002). Following that storm, large changes were observed in grain size at several MWRA monitoring stations (for example, station NF02 changed from >75% silt-clay in 1992 to less than 10% silt-clay for the remainder of baseline); these changes were associated with changes in the infaunal communities. Similar impacts from the 1992 storms on grain size were not observed at farfield stations, most of which are substantially deeper than those in the nearfield and all of which are in more extensive areas of depositional sediments.

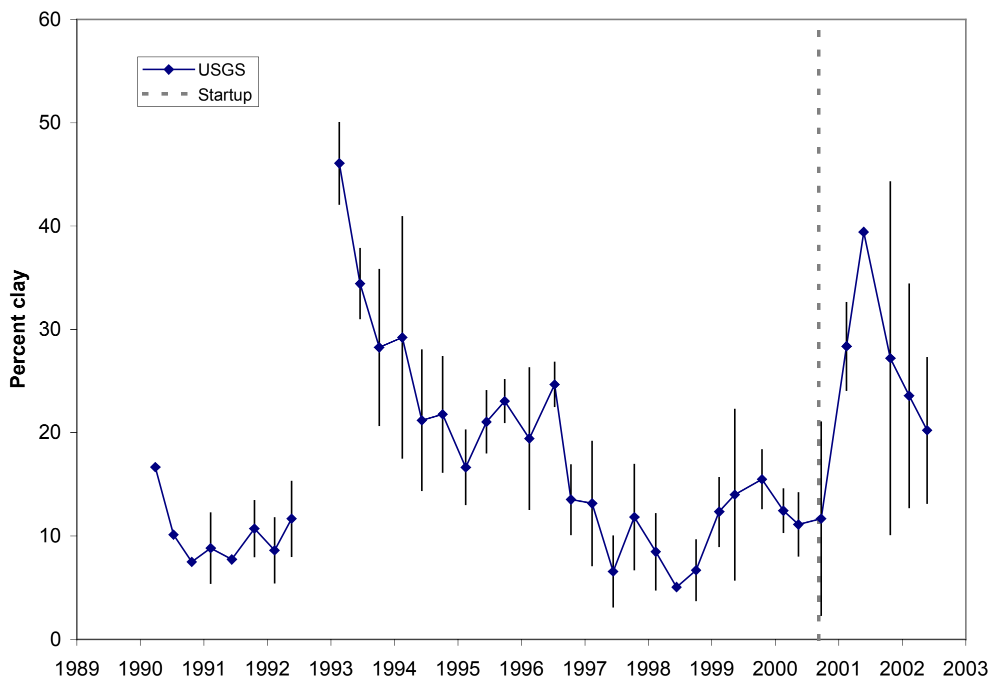


Figure 3-6 Percent clay in USGS surficial sediments from station NF12, 1989-2002.

3.3.2 Infaunal monitoring

Soft-bottom sediments in the nearfield support typical New England coastal benthic communities. Stations with fine sediments have communities dominated by polychaete worms, such as *Prionospio steenstrupi*, *Spio limicola*, *Mediomastus californiensis*, and *Aricidea catherinae*. Sandier stations are inhabited by polychaetes *Polygordius* sp. A and species in the genus *Exogone* and by amphipods such *Crassicorophium crassicorne* and *Unciola* spp. Communities in the nearfield through baseline were characteristic of New England shallow subtidal sediments subjected to natural disturbance (Hilbig and Blake, 2000), for example sporadic sediment resuspension and transport.

Infaunal benthic communities found at farfield stations share many species with those found in the nearfield. *P. steenstrupi* and other spionid polychaetes are often the numerical dominants in both nearfield and farfield sediments, as in 2002. Farfield stations also support a wide variety of other species characteristic of New England coastal habitats. Polychaete worms, including *Euchone incolor*, *Aricidea quadrilobata*, and *Levinsenia gracilis*, predominate at most stations. Benthic infaunal monitoring data since outfall startup in 2000 does not indicate major departures from the baseline monitoring period. Nearfield infaunal abundances in 2002 were higher than previously observed (Figure 3-7). This appears to be a result of increases in the abundance of several polychaete species normally present in nearfield muddy sediments, for example the small deposit feeding spionid *Prionospio steenstrupi* (Figure 3-8), the capitellid *Mediomastus californiensis,* and the cirratulid *Tharyx acutus*. *P. steenstrupi,* which has been one of the infaunal community dominants in the nearfield since the mid-1990s, had a nearfield mean abundance of 1,215/sample in 2002, compared to a maximum in previous years of 1,033 (1999). It is not clear whether the increased infaunal abundances observed in 2002 are a response to outfall discharge or not; this will continue be further investigated as data from 2003 become available.

In 2002 only one station (NF17) was not dominated by a polychaete species; at this station, the ascidian *Molgula manhattensis* accounted for over 25% of the infauna. This taxon was not consistently identified to species in previous years, because most animals observed were tiny juveniles. In 2002 some were large enough to identify to species, and were therefore combined with the juvenile ascidians recorded previously. Under this treatment of the data*, M. manhattensis* appears as the top dominant at NF17 in 2001, and ranked second in 2000. The species has been numerous at several other sandy stations in recent years, for example NF13, NF19, and NF23.

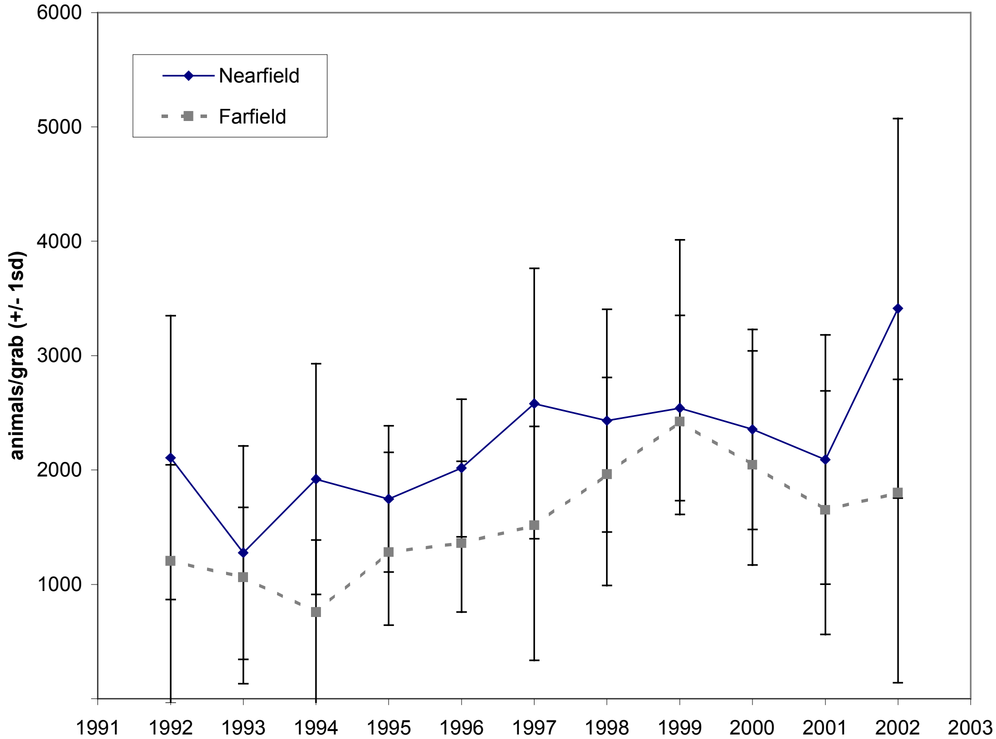


Figure 3-7 Average infaunal abundance in nearfield and farfield samples, 1992-2002.

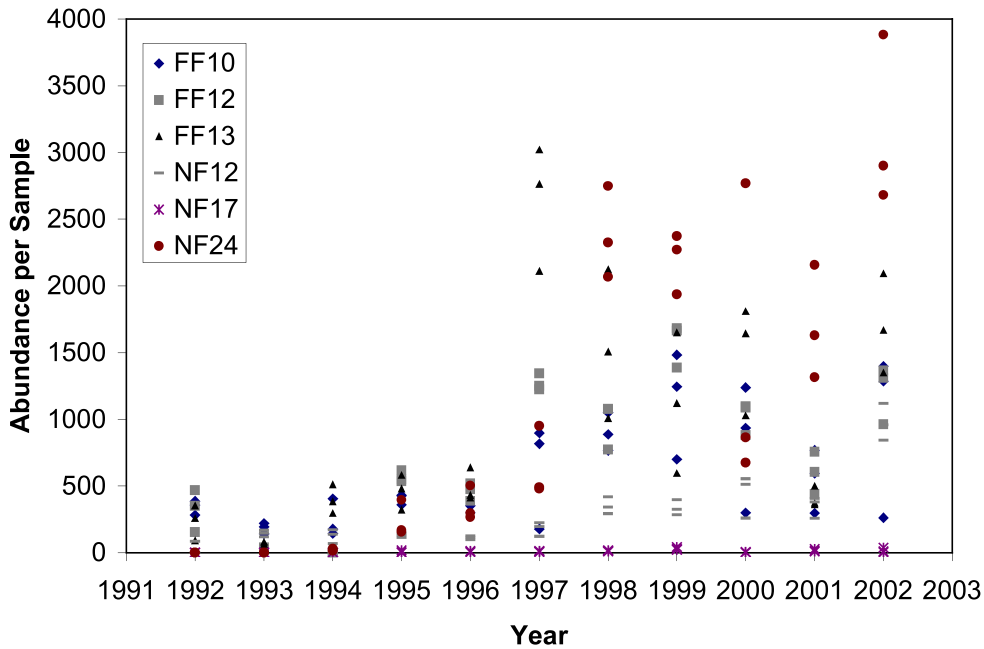


Figure 3-8 Abundance of the spionid polychaete *Prionospio steenstrupi* in samples from replicated nearfield stations, 1992-2002.

Multivariate community composition analyses support these observations. PCA-H analyses (Trueblood *et al.* 1994) of the combined, 640 sample nearfield and farfield datasets documented that these observations are reflected in quantitative analyses of the community composition data. Using a CNESS sample size of 15, the first 2 factors of the analysis explained 31% of the variability in the monitoring data. Fully labeled graphics including the 2002 data were not available in time for this briefing package, but the preliminary plots document that infaunal communities and trends observed during baseline continue. Figure 3-9 shows that regional infaunal communities are arrayed in an arc along what appear to be gradients in grain size and depth (or factors associated with them). Most of the separation along Axis 1 is associated with the distinction between the fine-grained mud communities observed in the nearfield (negative loading) and communities at the deepwater farfield stations off Cape Anne and in Stellwagen Basin (FF11, FF14, FF04, FF05), which have positive Axis 1 loadings. Samples from Cape Cod Bay (FF06, FF07) consistently plot between the nearfield and deepwater samples. Axis 2 from the PCA-H serves to separate communities along the gradient in grain size observed in the nearfield. The samples from sandy stations like NF17, NF14, NF04, and NF23 plot at one extreme of Axis 2, with positive loadings, while samples from muddier sites (e.g. NF12, NF08, NF24) have negative loadings on Axis 2.

The Gabriel Euclidian distance biplot (Gabriel, 1971) for this analysis shows the taxa whose relative abundances contribute significantly to this pattern (Figure 3-10). Taxa with relatively long vectors have stronger contributions to the variance in community structure. Acute angles between vectors indicate taxa which tend to co-occur in samples, while oblique angles indicate taxa that tend not to co-occur. These tend to be the abundant or dominant taxa for the different environments already described, or, like the polychaete *Cossura longocirrata*, are species routinely present in samples from one region while absent from other regions. *C. longocirrata* is often a community dominant at Cape Cod Bay stations (for example, it made up half of all animals counted at FF07 in 2002). It is also fairly ubiquitous in Stellwagen Basin (e.g. stations FF04, FF11), though rarely a community dominant, with abundances of dozens to hundreds per grab. In contrast, *C. longocirrata* is very rarely found in nearfield samples, and then usually at counts of less than 10 per sample.

We ran separate PCA-H and clustering analyses on nearfield and farfield data to allow us to better separate out infaunal assemblages in the 2 regions. In the nearfield analysis, 8 species are responsible for 42% of the variation in community structure (measured by CNESS, m = 15). These are (in rank order) the polychaetes *Spio limicola* (7%), *Prionospio steenstrupi* (7%), *Aricidea catherinae* (5%), *Mediomastus californiensis* (5%), *Dipolydora socialis* (5%), *Tharyx acutus* (5%), *exogone hebes* (4%), and the amphipod *Crassicorophium crassicorne* (4%). The cluster analysis revealed four nearfield species assemblages: 1) a *Spio limicola* - *Prionospio steenstrupi –Mediomastus californiensis* assemblage, 2) an *Aricidea catherinae –Tharyx acutus – owenia fusiformis* assemblage, 3) a *Diploydora socialis* assemblage, and 4) an *Exogone hebes-crassicorophium crassicorne* assemblage.

The farfield PCA-H analysis showed that six species are responsible for 40% of the variation in community structure (measured by CNESS, m=15). These are in rank order *Prionospio steenstrupi* (10%), *Spio limicola* (8%) *Cossura longocirrata* (7%) *Euchone incolor* (6%) *Dipolydora socialis* (5%), and *Chaetozone setosa* MB (5%). A cluster analysis identified four distinct species assemblages: 1) *P. steenstrupi*, 2) *S. limicola – Dipolydora*, 3) *Chaetozone setosa* MB (including *Aricidea quadrilobata*, and 4) A *Cossura longocirrata – Euchone incolor*.

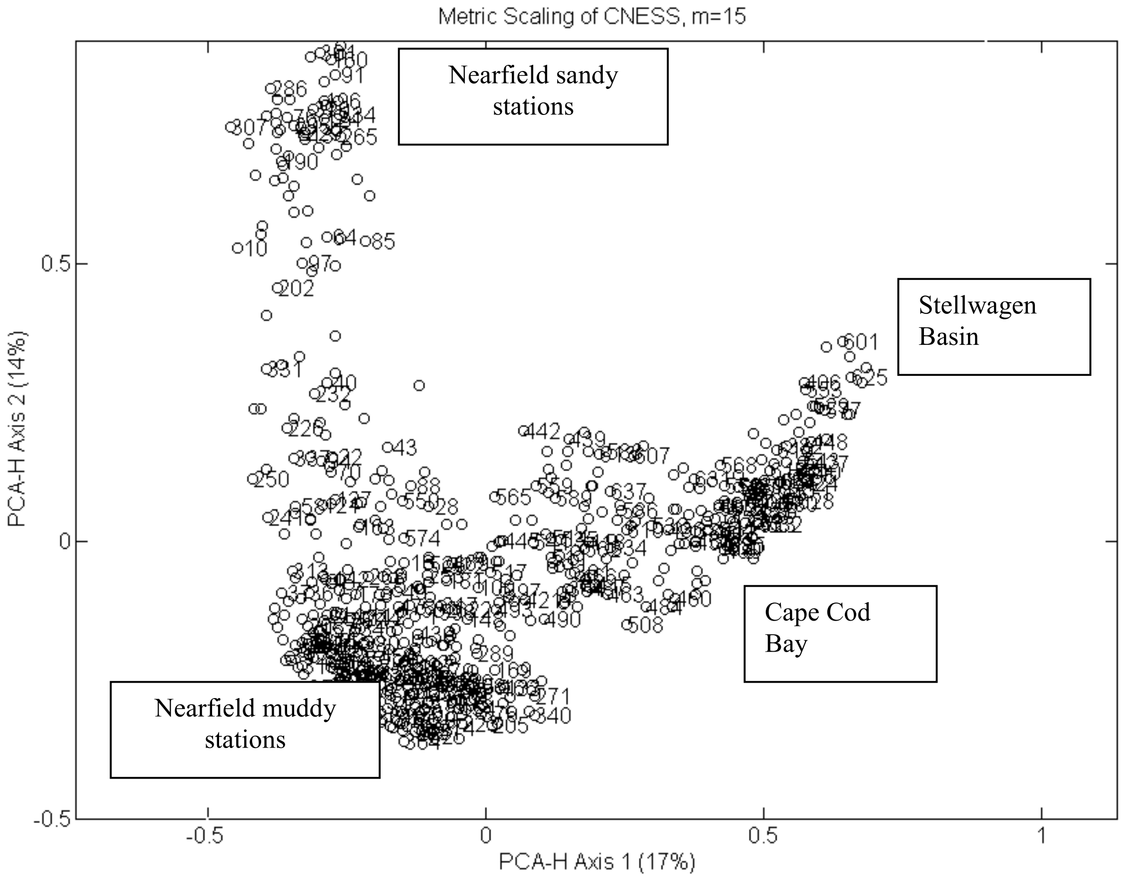


Figure 3-9 Metric scaling plot of CNESS distance, PCA-H Axis 1 versus Axis 2, among the 640 nearfield and farfield samples collected 1992-2002. Text boxes show regions whose samples consistently plot in that area of the graph.

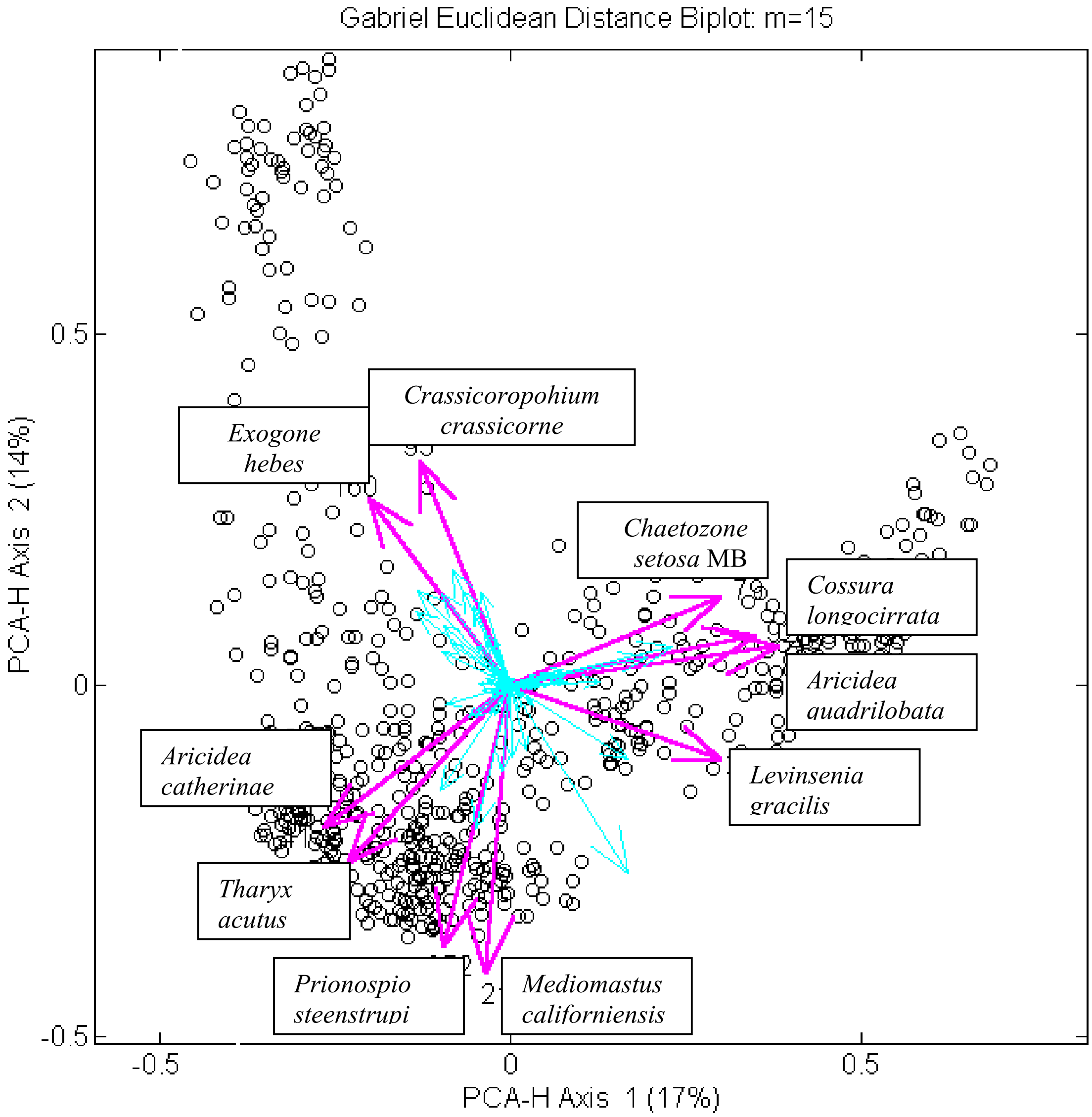


Figure 3-10 Gabriel Euclidian distance biplot , axes 1 versus 2, for the 640 nearfield and farfield infaunal samples from 1992-2002, showing those species that control the orientation of samples shown in figure 3-9. Vectors for species explaining the most variability in the plot are in dark gray (pink in color plots) with labels added; taxa explaining less variability are plotted in light gray unlabeled vectors (green in color plots).

3.3.3 Patterns of similarities between samples

In the PCA-H analysis, samples from 2001 and 2002 (discharge years) do not show any departure from the patterns exhibited by baseline data; that is, there are no indications of an outfall impact. That is not obvious in the available preliminary graphs from the PCA-H analyses (Figures 3-9, 3-10), but we have pursued an analysis of the patterns of between-sample similarity through time that conclusively demonstrates that there is no detectable impact of outfall discharge on community composition in either the nearfield or farfield.

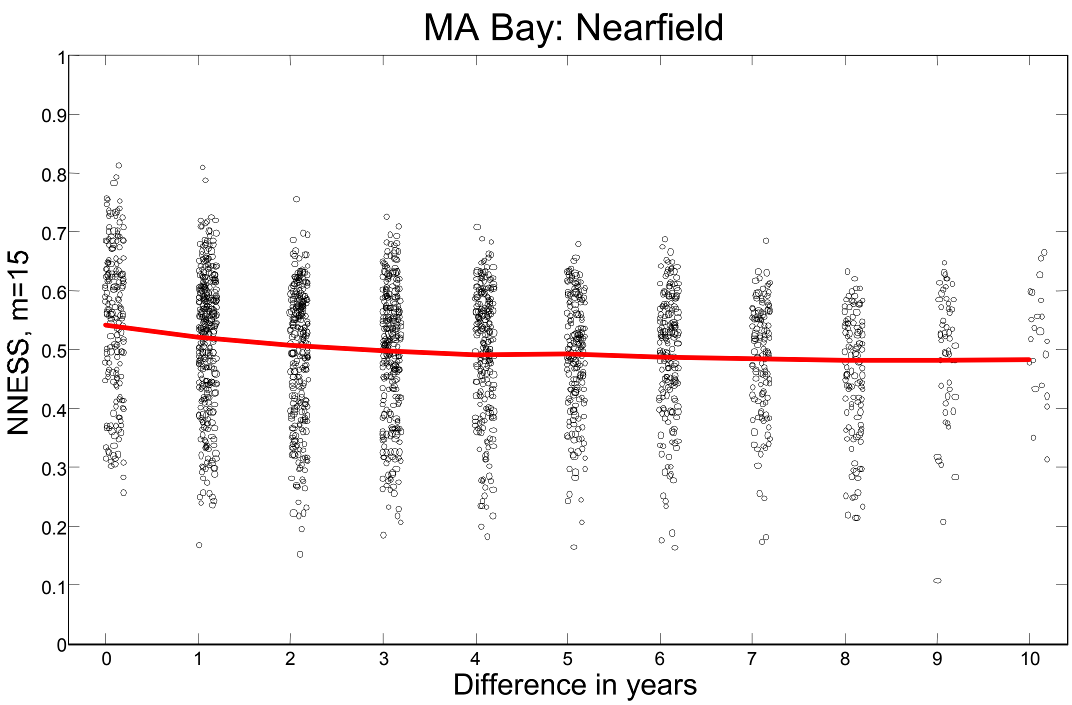


Figure 3-11 Time-sorted NNESS values (m = 15) for nearfield (top) and farfield(bottom) samples, 1992-2002, with best-fit LOESS lines.

In summary, decreases in between-sample similarity as pairs of samples get further apart in time are small and slow (Figure 3-11). Details of the evaluation can be found in Attachment A to this chapter. Importantly, there are no changes to these patterns, either in nearfield or farfield sediments, for pairs of samples between six and ten years apart. If there were changes in community composition associated with outfall startup, similarities between pairs of baseline and discharge samples would tend to be lower than similarities between pairs of discharge samples or pairs of baseline samples. At six years separation, approximately 40% of the NNESS similarities are between baseline and discharge samples. This increases to about half of similarities for samples separated by seven years, to two-thirds of all 8-year comparisons. Finally, all nine and ten-year comparisons involve similarities between baseline and discharge samples. The flat or even slightly increasing LOESS curves observed in all evaluations (Figure 3-11and Attachment A) at temporal separations of 6 to 10 years are strong evidence that there are no changes in infaunal community composition, either in the nearfield or farfield, coincident with outfall startup.

3.3.4 Diversity and abundance trends.

During the nine years of baseline monitoring, annual measurements of community parameters showed somewhat similar temporal patterns in the nearfield and farfield. In the nearfield, there was a large reduction in overall abundance and species richness of species between 1992 and 1993 (Figure 3-7, Figure 3-12). This decline has been attributed to the severe winter storm in December 1992. The effects of the storm were not apparent in the farfield. Infaunal abundance (Figure 3-7) increased in both the nearfield and farfield in the mid 1990s, with a baseline peak in 1999. As mentioned previously, nearfield abundance increased in August 2002 to levels not previously observed.

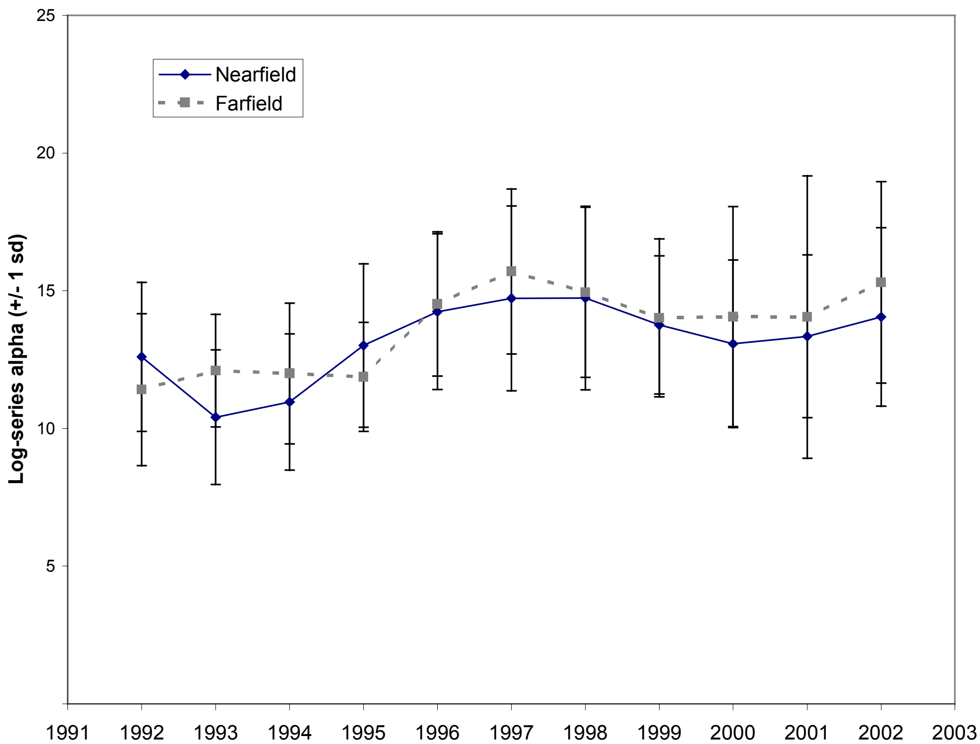


Figure 3-12 Average infaunal log-series alpha (a measure of species richness, relatively insensitive to changes in sample abundance) at nearfield and farfield stations, 1992-2002.

Species richness in both nearfield and farfield sediments increased in 1994 through 1997, decreased between 1998 and 2000, then increased in both regions in 2001 and 2002 (Figure 3-12). Overall, data for both nearfield and farfield stations are suggestive of a long-term cycle in species richness. This apparent pattern needs to be considered in evaluating discharge monitoring data. Changes in nearfield species richness not observed in the farfield could be indicative of an outfall impact.

We continue to evaluate the most appropriate statistical design for analyzing the monitoring data; this section describes results of analyses conducted through mid-July. We conducted a series of general linear model (GLM) analyses on diversity data from all 640 infaunal monitoring samples. Our GLM analyses are factorial ANOVA models, using station, year, near vs. far, and pre- and post as categorical explanatory variables. We worked with log-series alpha and with Pielou's J' to investigate whether there are significant differences in species richness or evenness between the nearfield and farfield, significant differences among years, whether any outfall impact can be detected and how sensitive the design is to changes in species richness that might result from the outfall. These analyses allow powerful tests for detecting discharge effects in the monitoring data.

Our design is a form of Green's optimal impact design. Green (1979) described an optimal impact design study as one in which baseline data exist adjacent to the site of a potential environmental impact and at a spatially distant control area, removed from the potential influence of the impact. Green proposed that an impact could be assessed through the use of a two-factor analysis of variance (ANOVA), but his design was criticized by Hurlbert (1984) who argued that replicated affected areas must be sampled. Green proposed monitoring just two areas and then testing for an area by post-impact interaction effect. Hurlbert rightly argued that all natural communities change through time and that natural differences among areas could be confused as an impact. A modified version of the optimal impact design study was proposed by Stewart-Oaten *et al.* (1986); their design had been used to assess the effects of nuclear power plant cooling water in near-shore California marine habitats. They never implemented the plan to detect effects. Underwood (1997) reviewed this class of ANOVA model, which have come to be known as Before-After-Comparative Impact (BACI) designs. He posed different types of change that might confound a test for impact, but to our knowledge, our design is one of the first that directly tests potential effects of the MWRA outfall. We control for station effects, and long-term patterns and detect outfall effects despite long-term cyclic patterns in several of the key response variables in different regions. Due to the long duration of our baseline and the relatively large sample sizes, our tests are quite powerful for detecting potential effects of the outfall.

Log-series alpha. The preliminary GLM analysis confirmed the pattern observed in Figure 3-12. The analytical model included year and region factors, and the year-region interaction. Neither the region nor the interaction terms in the model are significant in this analysis, see Table 3-2. The estimated marginal means for the significant year effect (Figure 3-13) emphasize the apparent 7-year cycle in species richness observed in the Bays.

Table 3-2 Results of GLM analyses on log-series alpha using year and region as factors.

Tests of Between-Subjects Effects
Dependent Variable: log series alpha

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Corrected Model	1295.06[b]	21	61.67	6.73	.000	141.27	1.000
Intercept	111782.26	1	111782.26	12193.20	.000	12193.20	1.000
NEAR	33.05	1	33.05	3.60	.058	3.60	.474
YEAR	1102.94	10	110.29	12.03	.000	120.31	1.000
NEAR * YEAR	118.05	10	11.80	1.29	.233	12.88	.670
Error	5702.24	622	9.17				
Total	121976.83	644					
Corrected Total	6997.30	643					

a Computed using alpha = .05
b R Squared = .185 (Adjusted R Squared = .158)

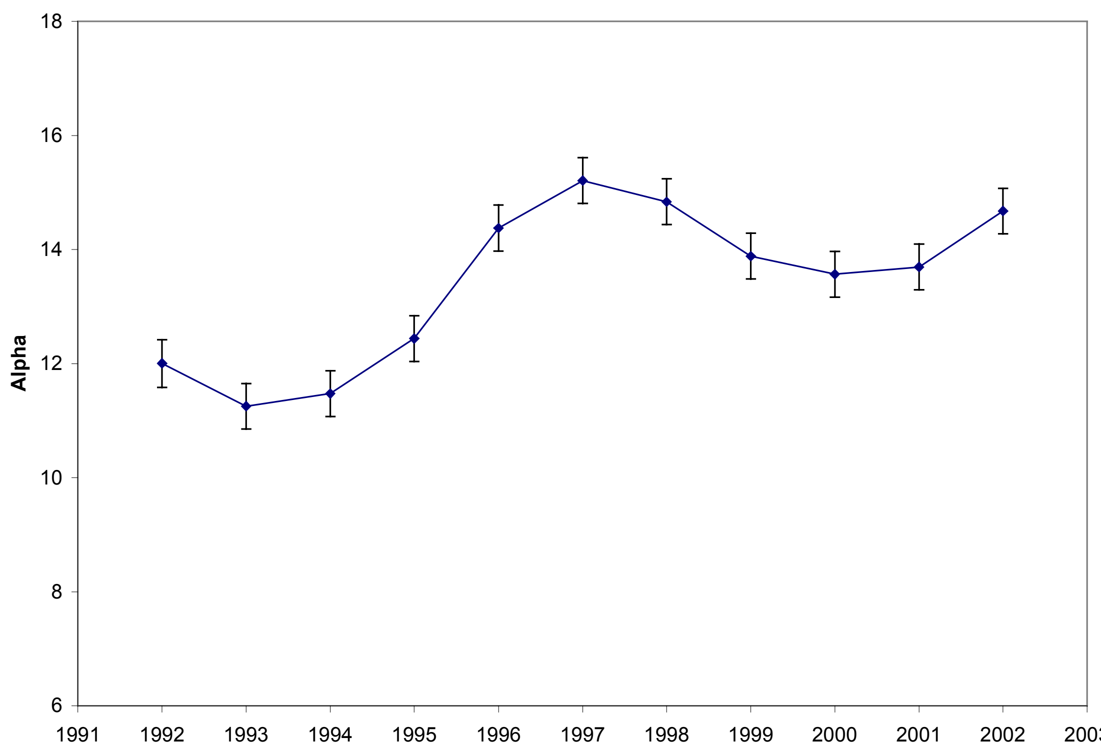


Figure 3-13 Estimated marginal means and standard errors for the year effect modeled in Table 3-2.

There was a consistent nonrandom pattern in the untransformed log-series alpha residuals after fitting the log-series alpha model. This problem was resolved by taking the natural logarithm of Fisher's alpha. While the test for non-equal variances is still significant, the departures from expectation and test statistic are relatively small ($F<2.0$), with the significance being driven by the 200-300 degrees of freedom in the numerator and denominator for the Levene's F test for equal variance among groups.

The second GLM model was designed as a sensitive test for possible outfall effects on species richness. Using Type I hierarchical decomposition of the sum of squares, this model included:

- year effects,
- station effects,
- the region-discharge interaction term (NEAR*PRE)
- the year*station interaction.

The ANOVA table (Table 3-3) contains the F statistic for NEAR*PRE to determine whether the difference in alpha between nearfield and farfield changed after the outfall went online in September 2000. The F statistic is 1.0, indicating no change in the difference in alpha between pre and post. An assessment of the estimated effect sizes indicate that this design could detect about a 6% change in Fisher's alpha due to the outfall. This is would be a very small effect, given that Fisher's alpha increased by over 45% between 1993 and 1997.

Table 3-3 Results of GLM analysis modeling year, station, region*discharge interaction, and year*station interaction terms using hierarchical decomposition of sum of squares.

Tests of Between-Subjects Effects

Dependent Variable: ln (alpha)

Source	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Obs. Power[a]
Corrected Model	35.8[b]	329	0.11	9.307	1.90E-74	0.907	3062.0	1
Intercept	4227.4	1	4227.45	361603.2	0	0.999	361603.2	1
YEAR	6.8	10	0.68	58.442	1.03E-65	0.650	584.4	1
STATION	19.2	35	0.55	46.932	4.92E-104	0.840	1642.6	1
NEAR * PRE	0.0	1	0.01	0.999	0.318	0.003	0.999	0.169
YEAR * ST ATION	9.8	283	0.03	2.947	2.03E-20	0.726	834.0	1
Error	3.7	314	0.01					
Total	4266.9	644						
Corrected Total	39.5	643						

a Computed using alpha = .05

b R Squared = .907 (Adjusted R Squared = .810)

Pielou's J' We analyzed the species evenness data using the same GLM model. This model suggests a significant outfall effect (Table 3-4), even though on average species evenness increased in the nearfield in 2001 and in 2002 was well within the range observed during baseline (Figure 3-14).

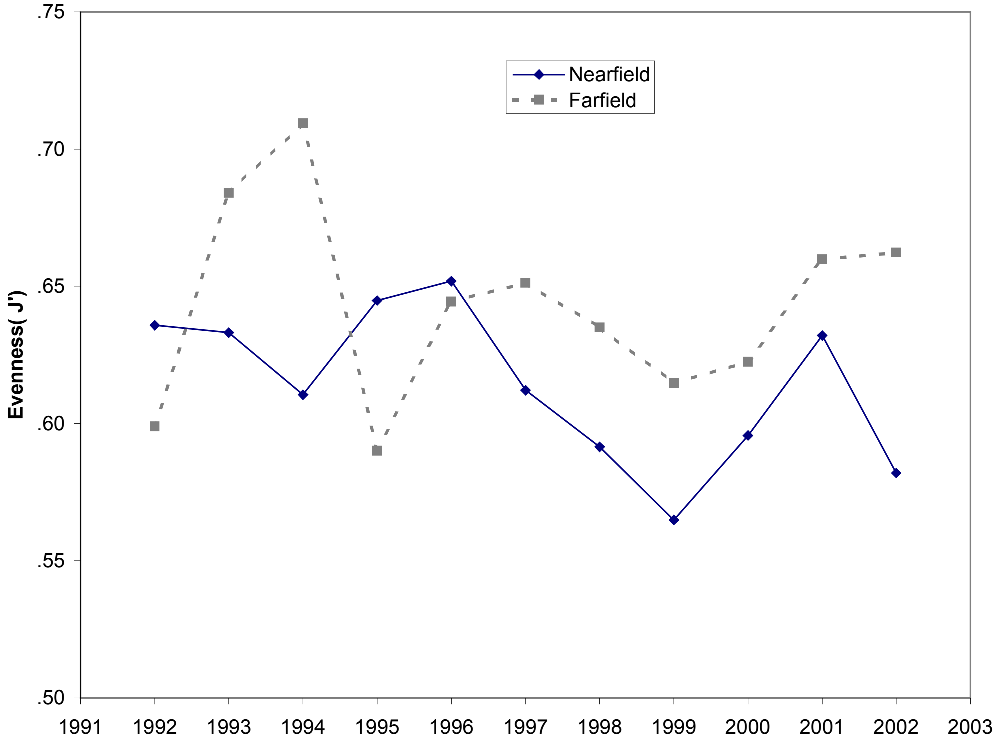


Figure 3-14 Average evenness (Pielou's J') in nearfield and farfield sediments, 1992-2002.

Table 3-4 Results of GLM analysis on Pielou's J', modeling year, station, region*discharge interaction, and year*station interaction terms using hierarchical decomposition of sum of squares.

Tests of Between-Subjects Effects

Dependent Variable: Pielou's J'

Source	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Obs. Power[a]
Corrected Model	6.09[b]	329	0.019	8.38	0.000	0.898	2758.4	1
Intercept	252.12	1	252.1	114232.3	0.000	0.997	114232.3	1
YEAR	0.28	10	0.028	12.62	0.000	0.287	126.2	1
STATION	2.83	35	0.081	36.58	0.000	0.803	1280.5	1
NEAR * PRE	0.03	1	0.031	14.17	0.000	0.043	14.2	0.964
YEAR * STATION	2.95	283	0.010	4.73	0.000	0.810	1337.6	1
Error	0.69	314	0.002					
Total	258.90	644						
Corrected Total	6.78	643						

a Computed using alpha = .05

b R Squared = .898 (Adjusted R Squared = .791)

Computationally, the GLM analysis fits a parallel lines model (no evidence for interaction) between near and far and then tests whether the difference in evenness between near and far is changed after the outfall went on line. The differences detected by the model are highly significant, but are based on a 0.03 difference in modeled J', compared to the pre-outfall baseline, (Figure 3-15). This is quite a subtle difference. Only through use of the GLM model including the station effect is the finding significant; models lacking that factor return a non-significant NEAR*PRE interaction term. Our use of the non-orthogonal hierarchic design produces a much more powerful analysis of possible treatment effects than would otherwise be possible, detecting a statistically significant result that does not appear to be ecologically important.

We have begun evaluating the change detected by the model. A preliminary examination of dominance-diversity curves, rarefaction curves, and fits to the log-series model suggests that the previously discussed increased abundance of community dominants observed in some nearfield samples in 2002 may account for the result. This issue will be further evaluated as the results of monitoring in 2003 and 2004 become available.

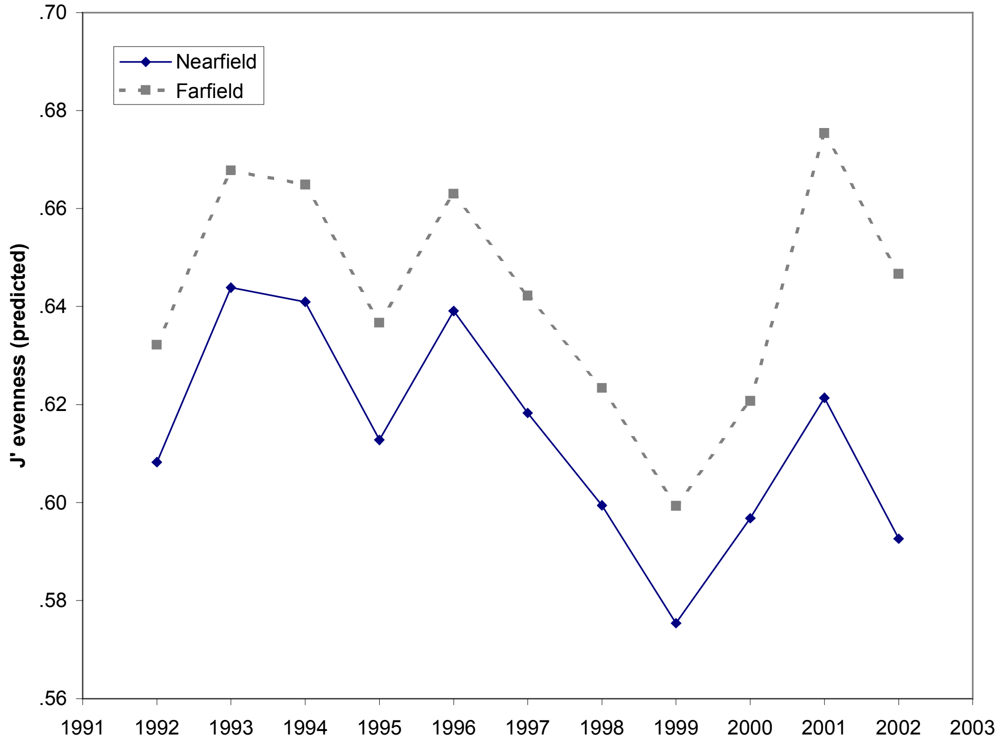


Figure 3-15 Predicted Pielou's evenness (J') for nearfield and farfield sediments, based on a GLM model including year, region, and the region*discharge interaction. The modeled separation between nearfield and farfield evenness increases by 0.03(from 0.024 to 0.054) following 2000.

3.3.5 Contaminant-infaunal interactions

We plan to address the monitoring question, "Are any benthic community changes correlated with changes in levels of toxic contaminants (or sewage tracers) in sediments" by conducting a redundancy analysis (Legendre and Legendre, 1998). Results of that analysis may not be available in time for the OMSAP workshop. However, it is important to note that, as we documented in the briefing for the April 1 OMSAP workshop (MWRA 2003A), there have been no indications of increases in contaminants in sediments as a result of outfall discharge. Similarly, and as documented in this briefing package, there have as yet been no meaningful changes in benthic infaunal communities attributable to the outfall. Therefore, while the redundancy analysis will help elucidate relationships between infaunal community composition, grain size, TOC, and concentrations of toxic contaminants in our monitoring dataset, there are few if any outfall effects for it to detect.

3.4 Future monitoring

MWRA proposes changes to infaunal monitoring to take effect in 2004 that are cost-effective and protective:

1. Continue the existing SPI study unchanged, collecting data annually from the 23 nearfield stations shown in Figure 3-2.
2. Continue to collect and analyze infaunal samples from stations NF12 and NF17 every year through 2005.
3. Randomly split the remaining stations into 2 subsets (Table 3-5). Collect and analyze infaunal samples from each subset in alternate years[2], so that all stations are sampled every two years. Retain current patterns of replication in infaunal and grain size samples. Stations where 3 infaunal and 2 grain size samples are collected will continue to have replicates collected, while unreplicated stations would remain unreplicated .
4. Review the benthic infaunal monitoring design during evaluation of the 2005 data to determine whether this design should continue or if additional modifications are appropriate. If no further changes are warranted, this design would continue.

OMSAP has previously approved MWRA's proposal to decrease the frequency of sediment contaminant sampling at most stations to every third year. If this proposal for benthic infaunal monitoring is adopted, MWRA will collect and analyze contaminant samples in 2005 at NF12 and NF17, and at all infaunal stations scheduled for sampling in 2005 (Table 3-5).

Table 3-5 Proposed subsets of nearfield and farfield stations for sampling in 2004 and 2005.

Year	Station Type	Stations	N
2004	Replicated nearfield	NF12, NF17, FF10, FF13	4
	Unreplicated nearfield	NF05, NF07, NF08, NF09, NF16, NF18, NF19, NF22, NF23	9
	Farfield (replicated)	FF04, FF05, FF07, FF09	4
2005	Replicated nearfield	NF12, NF17, FF12, NF24	4
	Unreplicated nearfield	NF02, NF04, NF10, NF13, NF14, NF15, NF20, NF21	8
	Farfield (replicated)	FF01A, FF06, FF11, FF14	4

[2] Stations were binned by region and level of replication before the random selection .

3.5 Justification

3.5.1 SPI sampling

Continued annual sampling at the 23 stations in or adjacent to the nearfield using sediment profile imaging (SPI) will allow us to rapidly determine if major changes have occurred in nearfield sediments (e.g. massive sediment transport events or the sediment RPD becmes shallower). This information will help MWRA evaluate the results from the sediment sampling. Given the one-week turnaround for preliminary results from the SPI survey, in an extreme case (e.g. widespread anoxic sediments) sediment sampling could be expanded that year to include all stations, rather than the subset scheduled. These data (which like the community composition data have shown no detectable change resulting from outfall discharge), will continue to be used to test MWRA's Contingency Plan threshold for sediment RPD **(**Figure 3-16).

3.5.2 Infaunal sampling

Replication The existing "hybrid" design for nearfield sampling, with some replicated and many more unreplicated stations, evolved in order to allow MWRA to make more powerful statistical inferences about the results (data from replicated stations), while obtaining quantitative community composition information at essentially all soft-sediment environments in the nearfield area (unreplicated stations). After careful consideration of the infaunal monitoring results to date, including the GLM analyses reported in Section 3.4, MWRA believes it is appropriate to continue this hybrid design in the nearfield, and to continue to collect replicate samples at farfield stations, although with reduced frequency at most stations.

Annual stations. Our rationale for proposing to continue annual infaunal community samples from stations NF17 and NF12 is consistent with the rationale for continuing to sample them for contaminants (MWRA 2003b). They are two of the most stable stations we sample for all parameters, including species composition. Samples from NF17 consistently reflect the sand community depicted in the upper left quadrant of Figure 3-9, while the fauna at station NF12 always reflects the nearfield mud community seen in the lower left of that figure.

Aditionally, patterns in species richness at these two stations is representative of that for the nearfield as a whole, (Figure 3-16). Access to detailed geophysical data from USGS means we would be better able to evaluate changes there than anywhere else, and, since MWRA will continue to sample NF12 and NF17 for contaminants annually under the proposals already approved by OMSAP, sampling for infauna at the same sites will allow an evaluation of whether changes in sediment contaminants at those sites (if observed) are associated with changes in the infauna.

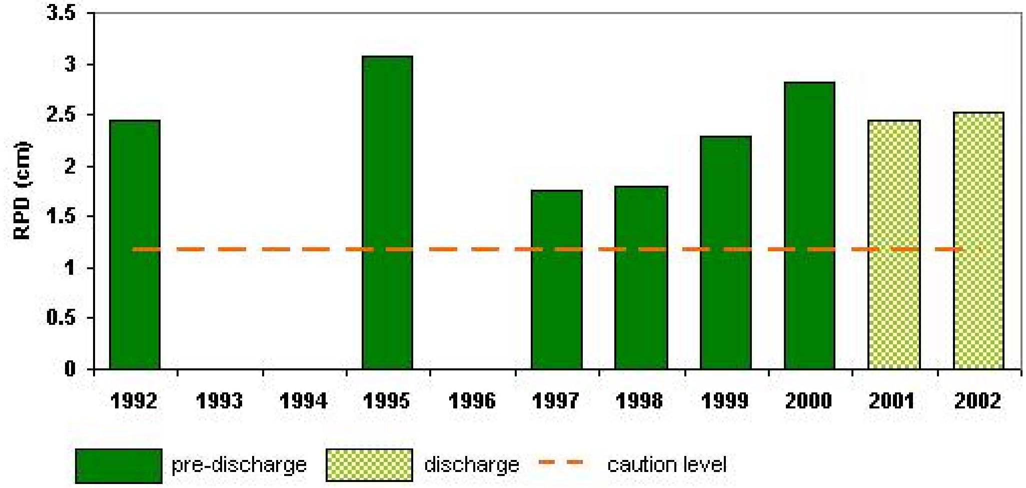


Figure 3-16 Mean Apparent Redox Potential Discontinuity (RPD) for nearfield sediments, 1992-2002, as measured by sediment profile imaging.

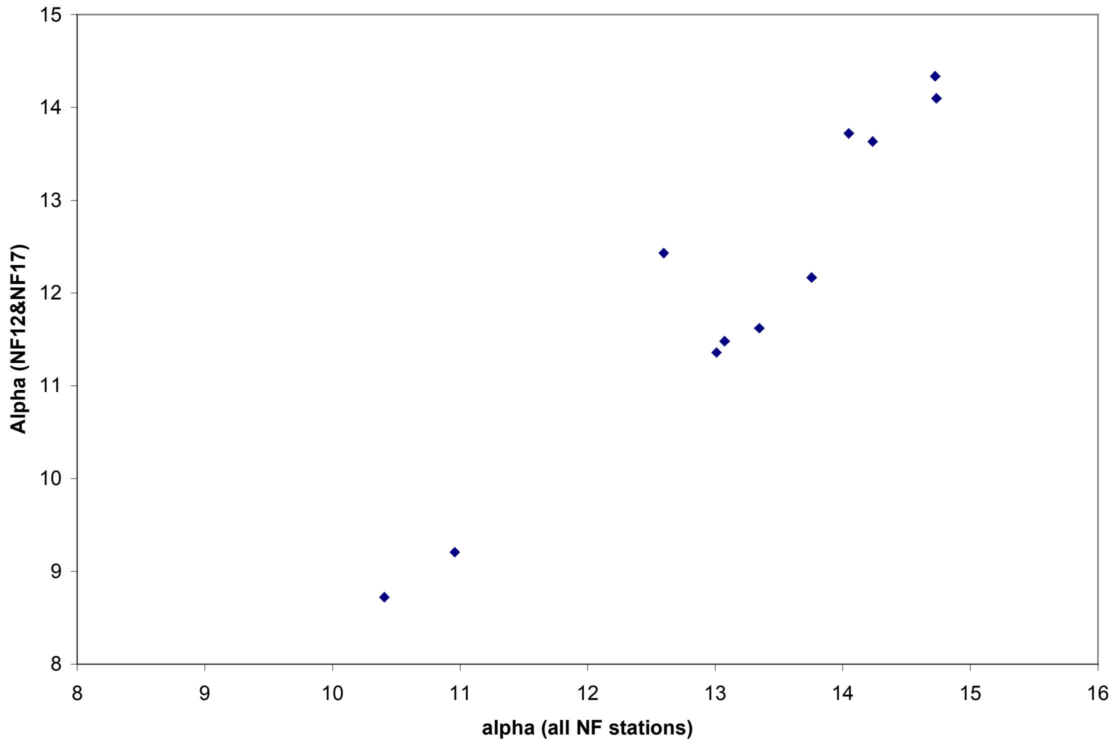


Figure 3-17 Biplot comparing mean species richness (as log-series alpha) at all nearfield stations to mean richness at stations NF12 and NF17, 1992-2002.

Alternating station subsets. The project team evaluated whether to put forth permanent station reductions at this time or a modified design relying on alternating subsets (as is being proposed). We ultimately decided that until additional years of discharge monitoring data are available, making large permanent reductions in the stations occupied (for example, permanently dropping half the farfield stations) is not yet appropriate. However, the evaluations described in this briefing package document that changes in infaunal community composition with time are both small and slow (Figure 3-11, Attachment A) and that species richness in the Bays undergoes long-term systematic changes (Figure 3-12). The station bins shown in Table 3-5 generate very similar patterns of time-sorted NNESS values and LOESS fits to each other and to the entire dataset, (Figure 3-17), and reproduce patterns in infaunal species richness observed in the whole dataset fairly well (Figure 3-18). Thus, we feel confident that the changes proposed here will continue to track natural fluctuations in infaunal community structure, while remaining sensitive to any possible impacts from the outfall discharge.

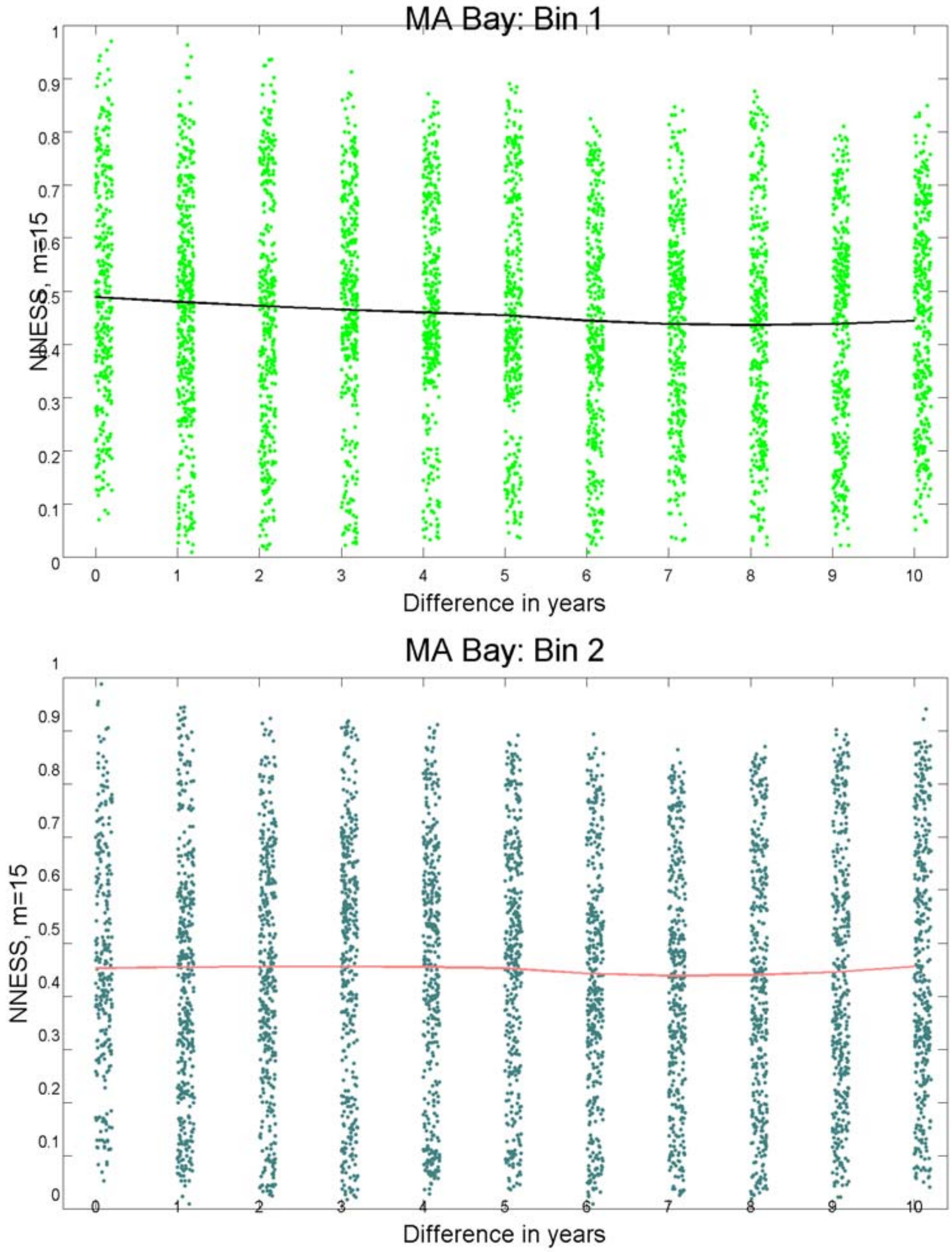
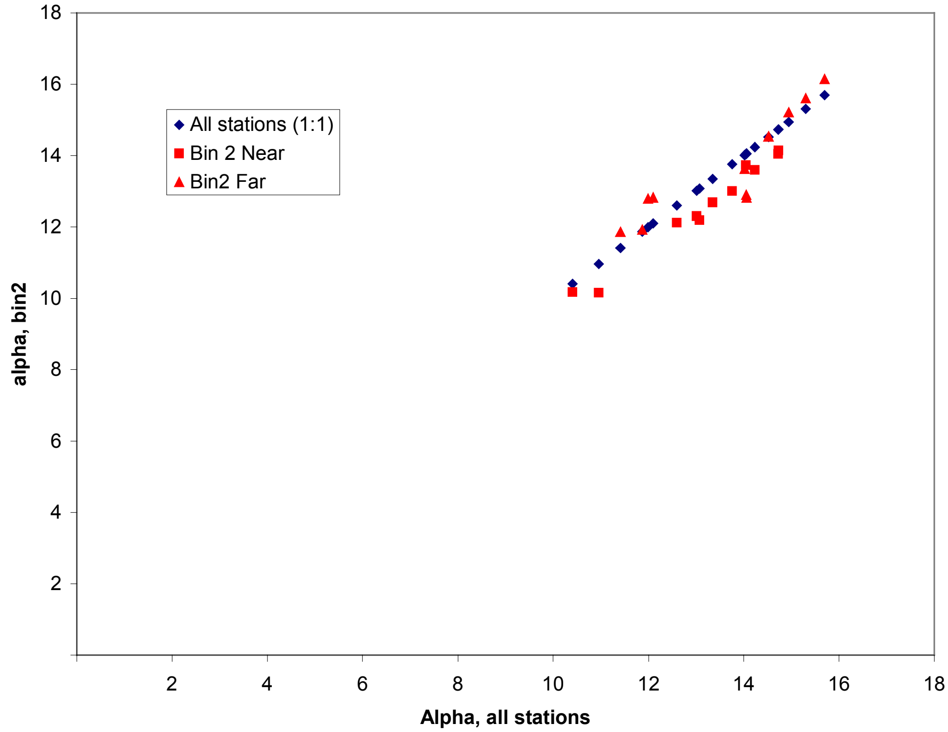


Figure 3-18 Time-sorted NNESS values (m = 15) for combined nearfield and farfield data 1992-2002 for the 2 proposed station bins, with best-fit LOESS lines.



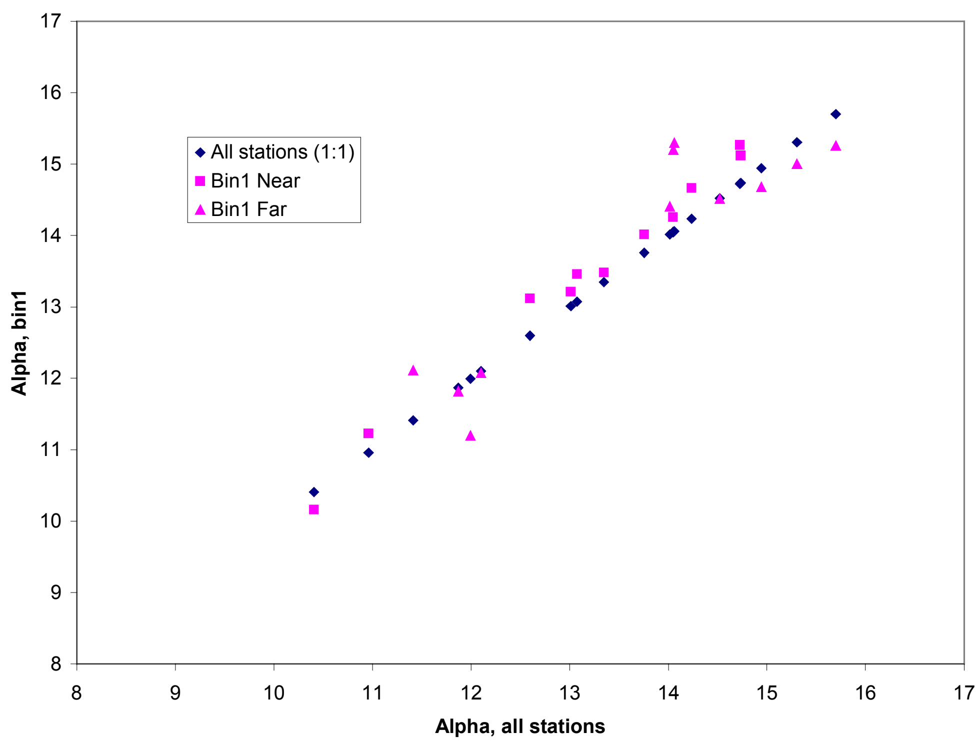


Figure 3-19 Biplot comparing 1992-2002 means for log-series alpha in nearfield and farfield sediments for bin1(top) and bin2(bottom) to those from all stations.

References

Blake JA, Hilbig B and Rhoads DC. 1993. Massachusetts Bay outfall monitoring program: soft bottom benthic biology and sedimentology, 1992 baseline conditions in Massachusetts and Cape Cod Bays. Boston: Massachusetts Water Resources Authority. Report 1993-10. 151 p.

Blake JA, Williams IP, Gallagher ED, Hecker B, Rhoads DC and Arnofsy PL. 1998. Massachusetts Bay outfall monitoring program: benthic biology and sedimentology baseline monitoring for 1997 and retrospective analysis of the 1992-1997 database. Boston: Massachusetts Water Resources Authority. Report 1998-16. 221 p.

Bothner, MH, Buchholtz ten Brink, M, and Manheim, FT, 1998, Metal concentrations in surface sediments of Boston Harbor - changes with time: Marine Environmental Research, v. 45, No. 2, p. 127-155.

Bothner, MH, Casso MA, Rendigs RR, and Lamothe PJ. 2002. The effect of the new Massachusetts Bay sewer outfall on the concentrations of metals and bacterial spores in nearby bottom and suspended sediments. Marine Pollution Bulletin v. 44, No. 2, p1063-1070.

Butman B, Bothner MH, Lightsom FL, Gutierrez BT, Alexander PS, Martini MA, and Strahle WS. 2002. Long-term Oceanographic Observations in Western Massachusetts Bay Offshore of Boston, Massachusetts: Data Report for 1989-2000. U.S. Geological Survey Digital Data Series DDS-74. DVD-ROM

Coats DA, Imamura E and Campbell JF. 1995. 1993 annual Soft-Bottom Benthic Monitoring: Massachusetts Bay Outfall Studies. Boston: Massachusetts Water Resources Authority. Report 1995-02. 137 p.

EPA. 1988. Boston Harbor Wastewater Conveyance System. Supplemental Environmental Impact Statement (SEIS). Boston: Environmental Protection Agency Region 1.

Gabriel,KR. 1971. The biplot graphic display of matrices with application to principal component analysis. Biometrika 58: 453-467

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. John Wiley, New York.

Hilbig, B. And J.A. Blake. 2000. Long-term Analysis of polychaete-dominated benthic infaunal communities in Massachusetts Bay. Bulletin of Marine Science. 67:147-164.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54: 187-211.

HydroQual and Normandeau. 1995. A water quality model for Massachusetts and Cape Cod Bays: Calibration of the Bays Eutrophication Model (BEM). Boston: Massachusetts Water Resources Authority. Report 1995-08. 402 p.

Knebel, H.J., and Circe', R.C., 1995, Seafloor environments within the Boston Harbor-Massachusetts Bay sedimentary system: a regional synthesis: Journal of Coastal Research, v. 11, p. 230-251.

Kropp RK, Diaz R, Hecker B, Dahlen D, Boyle JD, Abramson SL and Emsbo-Mattingly S. 2001. 2000 outfall benthic monitoring report. Boston: Massachusetts Water Resources Authority. Report 2001-11. 148 p.

A. ATTACHMENT A TO CHAPTER 3

A.1 Evaluation of between-sample infaunal similarities through time

Eugene D. Gallagher and Kenneth E. Keay

As one tool for investigating whether there have been appreciable changes in infaunal community composition during baseline or since outfall startup, we investigated temporal patterns in the between sample similarities for the monitoring dataset. This involved coupling a matrix of similarities between infaunal samples with a similar matrix of the years between pairs of samples, then fitting a line to the resulting biplot to determine if and how rapidly similarity decreases as the temporal distance between samples increases.

A.2 Methods

Similarity index NNESS is a similarity measure related to NESS (Smith & Grassle, 1977) modified to properly analyze samples containing singleton species (Trueblood *et al*. 1994). We used it in preference to CNESS distances because NNESS similarities can be fit to a decaying exponential, a pattern we considered could be displayed by the data. Similarity matrices were generated for the nearfield and farfield monitoring data at three NNESS "m" sampling sizes:

- m = 1. At a sample size of 1, NNESS similarities are weighted heavily towards the contribution of dominant species; subdominant or rare taxa contribute almost nothing to the computed similarities. At m = 1, NNESS is equivalent to the Morisita-Horn index (Magurran, 1988).
- m = 15. This NNESS sample size was determined to be the optimal trade-off between sensitivity of the index to the contribution of both dominant and relatively rare taxa in the dataset.
- m = infinity. For this analysis, NNESS was calculated on a presence/absence species/sample matrix instead of a matrix of sampling probabilities for a given "m". In this evaluation, NNESS similarities are least sensitive to changes in abundance of dominant taxa and maximally sensitive to the contributions of rare species, converging on similarities generated by Sorenson's presence-absence similarity index.

A.3 Time-sorted similarities

After the similarity matrices were generated, a similar matrix of temporal distance between samples (years) was constructed. The matrices were aligned, and a biplot of faunal similarities between pairs of samples as a function of the number of years between them was constructed. An illustrative example is shown in Table A-1 and Figure A-1 for a hypothetical design in which three stations are occupied in each of three consecutive years. Point "a" on the figure represents the faunal similarity (0.781) between samples from station 2 and station 3 in year three, which were 0 years apart. Point “b” graphs the similarity (0.392) between the sample from station 1 in year 1 and that from station 3 in year 3, two years apart.

The number of points on such a plot equals [n*n-1)]/2, where n is the number of samples. For the example, this generates two matrices with 36 values each [9*8)/2], which are easily aligned, plotted and visualized. For the 446 samples collected in MWRA’s nearfield monitoring 1992-2002, a full biplot would contain 99,235 pairs, five times as many as for a full biplot of the 198 samples collected thus far in farfield monitoring. To cut down on computational time and allow comparisons between plots of nearfield and farfield data not influenced by the larger number of nearfield samples, each time-sorted NNESS matrix was randomly subsampled to provide 4,000 points for the biplots. X-axis values (years between samples) were spread randomly across 25% of a year to allow better visualization of the patterns in similarity. For example, similarities for sample pairs one year apart were randomly assigned a separation ranging from 1 to 1.25 years. We used locally weighted regression (LOESS)(Cleveland 1993) to find the best fit lines through the points. The slopes of the LOESS fits give an indication of whether and how rapidly species composition may be changing through time. Since all nine and ten-year comparisons involve comparisons between baseline data and samples collected after outfall startup in 2001 and 2002, the slope of the LOESS fits in this area of the plots would indicate if compositional shifts had occurred in either the nearfield or farfield following outfall startup.

Table A-1 Similarity matrix (a) and matrix of temporal distance (b) (years between samples) for a hypothetical 3 station study sampled in 3 consecutive years. Bolded values are labeled in Figure A-1.

a)

Year	**Station**	**Faunal similarity matrix**								
1	1									
1	2	0.92								
1	3	0.945	0.981							
2	1	0.491	0.5	0.494						
2	2	0.543	0.566	0.548	0.907					
2	3	0.583	0.583	0.581	0.955	0.955				
3	1	0.573	0.563	0.574	0.914	0.967	0.955			
3	2	0.597	0.491	0.496	0.10	0.154	0.178	0.136		
3	3	0.392	0.424	0.421	0.06	0.117	0.142	0.101	0.781	
Year		1	1	1	2	2	2	3	3	3
Station		1	2	3	1	2	3	1	2	3

b)

Year	Station	Temporal distance matrix								
1	1									
1	2	0								
1	3	0	0							
2	1	1	1	1						
2	2	1	1	1	0					
2	3	1	1	1	0	0				
3	1	2	2	2	1	1	1			
3	2	2	2	2	1	1	1	0		
3	3	2	2	2	1	1	1	0	0	
Year		1	1	1	2	2	2	3	3	3
Station		1	2	3	1	2	3	1	2	3

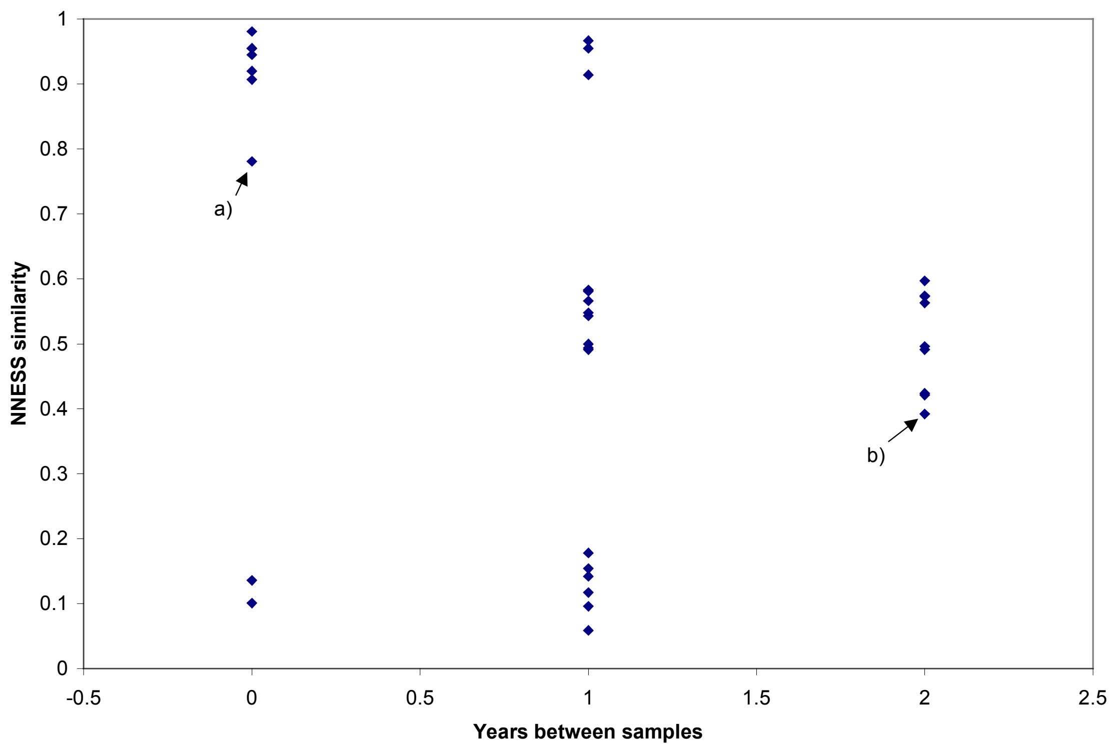


Figure A-1 Biplot of time-sorted NNESS values from Table A-1.

A.4 Results

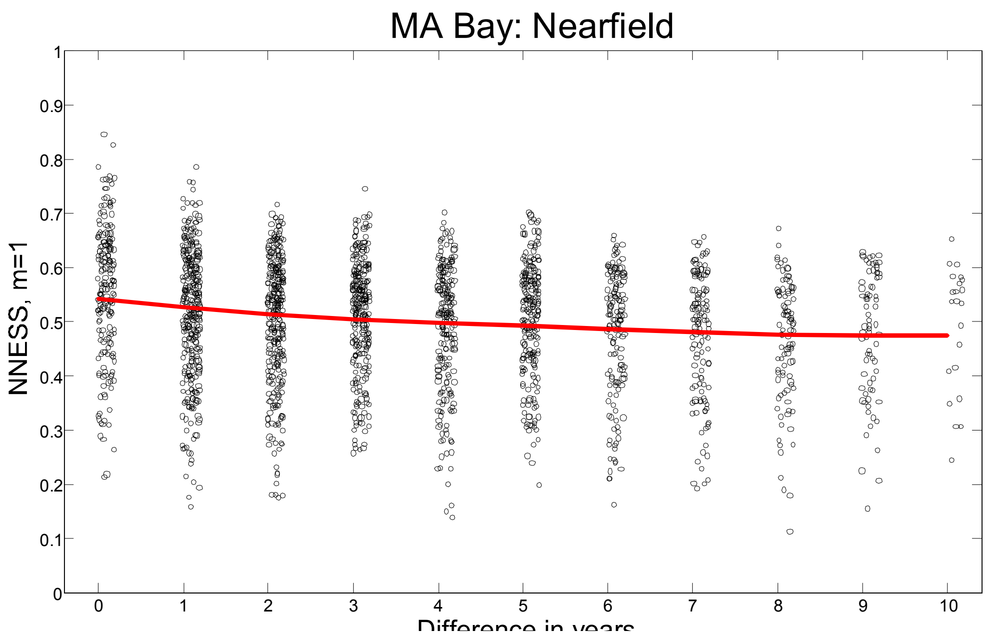
Figures A-2 through A-4 document that community composition in Massachusetts Bay samples (both nearfield and farfield) changes very slowly through time. The most rapid change in similarities are observed with NNESS at m=1 (Figure A-2). The slope of the LOESS fit for the farfield data in Figure A-2 corresponds to a roughly 20-year successional half-life. However, at m = 1, NNESS similarities are more sensitive to changes in relative abundance of community dominants than they are to changes in composition of rarer taxa. As the NNESS sample size ("m") increases and the similarity calculations become more sensitive to contributions from less abundant taxa (Figures A-3 and A-4), the LOESS slopes become flatter, indicating little if any change in community composition over decadal timescales in either the nearfield or the farfield.

Superimposing the nearfield and farfield biplots allows us to determine if the LOESS fits differ substantially between regions or if the slope of the LOESS fits is steeper for nearfield as opposed to farfield data. As can be seen for NNESS at m = 1 (figure A-5) , the nearfield and farfield LOESS fits are very similar.

As mentioned, all of the 9 and 10-year comparisons contrast similarities between pairs baseline and discharge samples data. In addition to that, approximately 40% of all six-year contrasts involve similarities between baseline and discharge samples, as do about half of those at 7 years separation, and roughly 2/3 of all samples at eight-year's separation are baseline/discharge comparisons. For all plots, at all NNESS sample sizes for both regions, LOESS fits are either flat or even suggest modest increases in between-sample similarity at seven to ten years' distance between samples. That is, there is no evidence in any of the plots for changes in community composition following outfall startup.

A.4.1 Impacts of proposed design changes

We carried out a similar analysis after splitting the 1992-2002 data into the station bins proposed for sampling in 2004 and 2005. For this evaluation, nearfield and farfield data were combined, and biplots of time-sorted NNESS similarities were constructed for the 2 station bins (Figures A-6through A-8). The two station bins produce patterns of changes in community composition through time similar to each other and to patterns observed in the entire dataset.



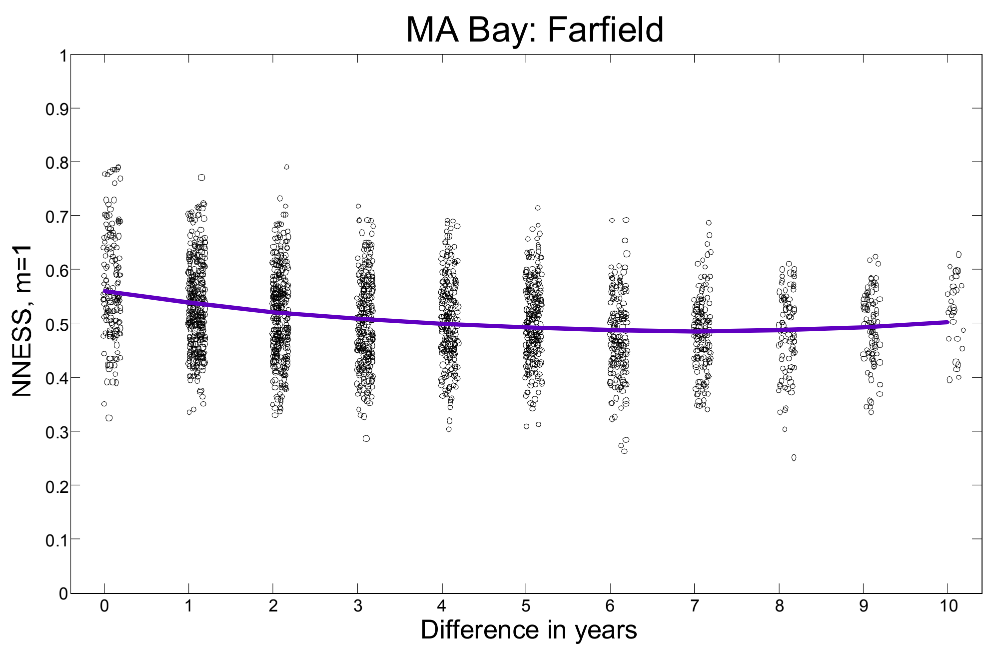
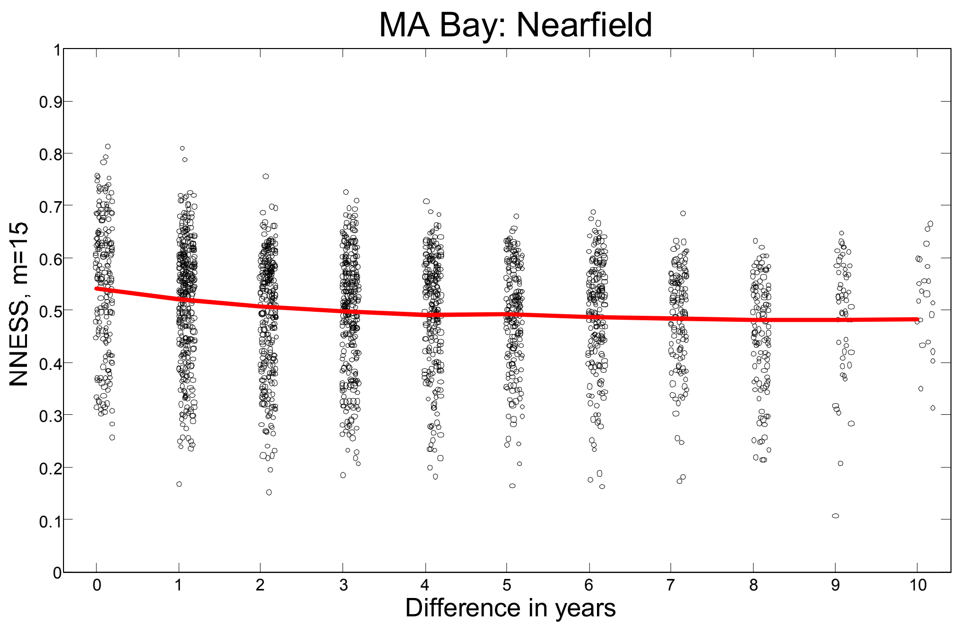


Figure A-2 Time-sorted NNESS values (m = 1) for nearfield (top) and farfield(bottom) samples, 1992-2002, with best-fit LOESS lines.



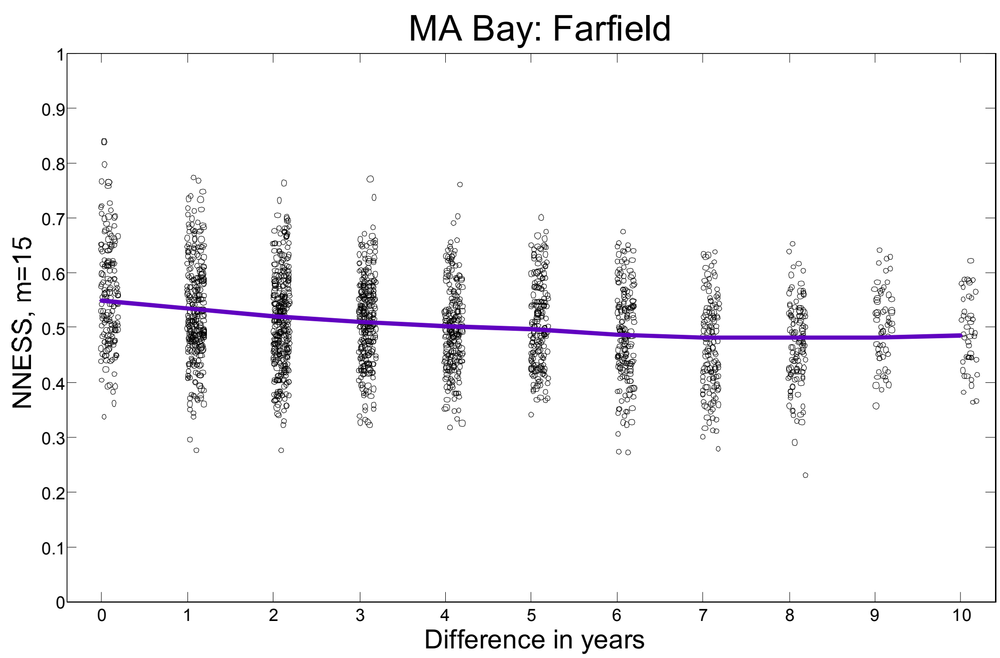
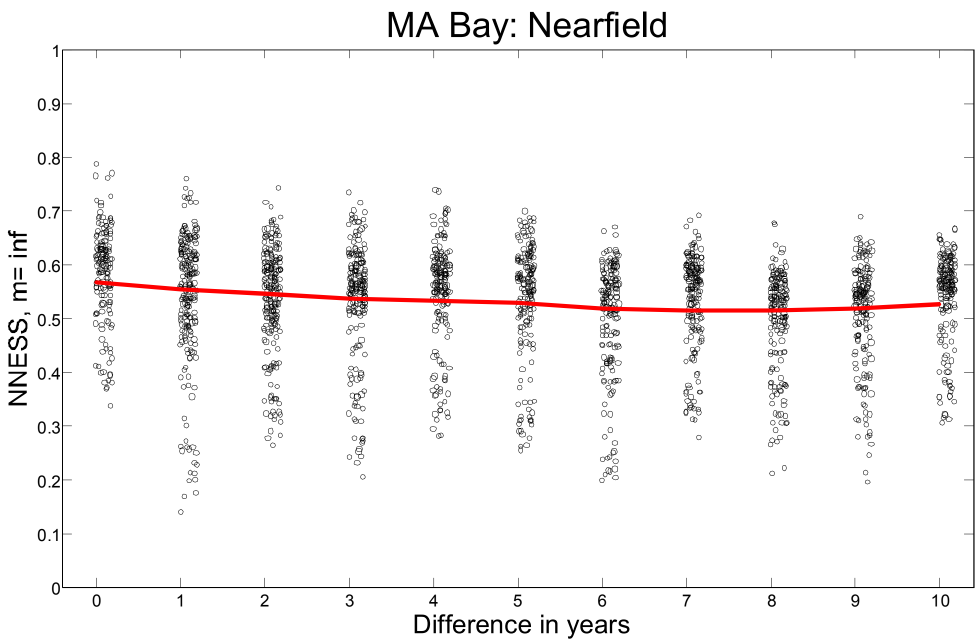


Figure A-3 Time-sorted NNESS values (m = 15) for nearfield (top) and farfield(bottom) samples, 1992-2002, with best-fit LOESS lines.



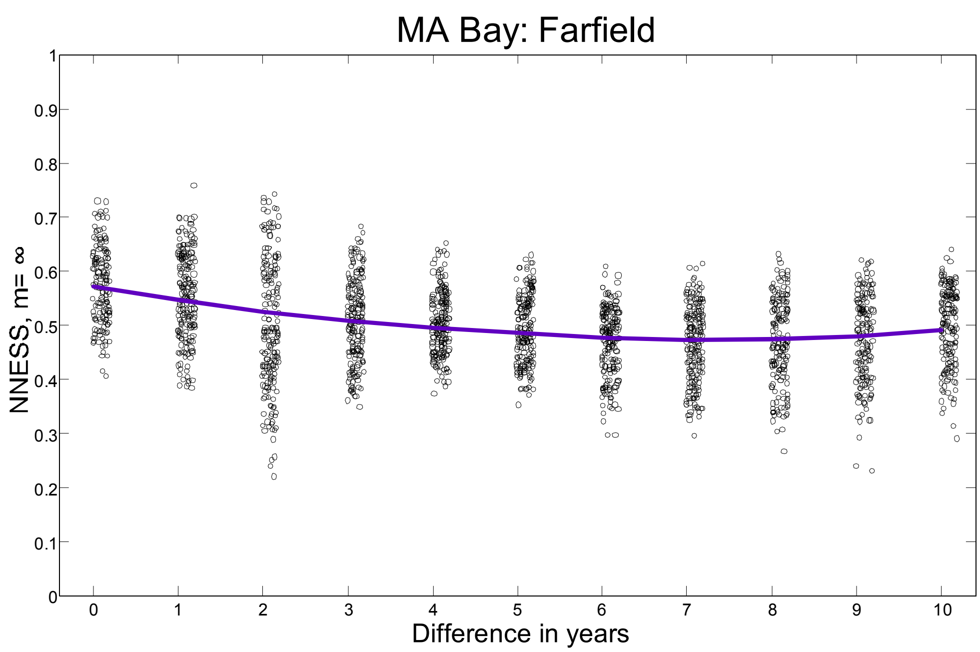


Figure A-4 Time-sorted NNESS values (presence-absence matrix) for nearfield (top) and farfield(bottom) samples, 1992-2002, with best-fit LOESS lines.

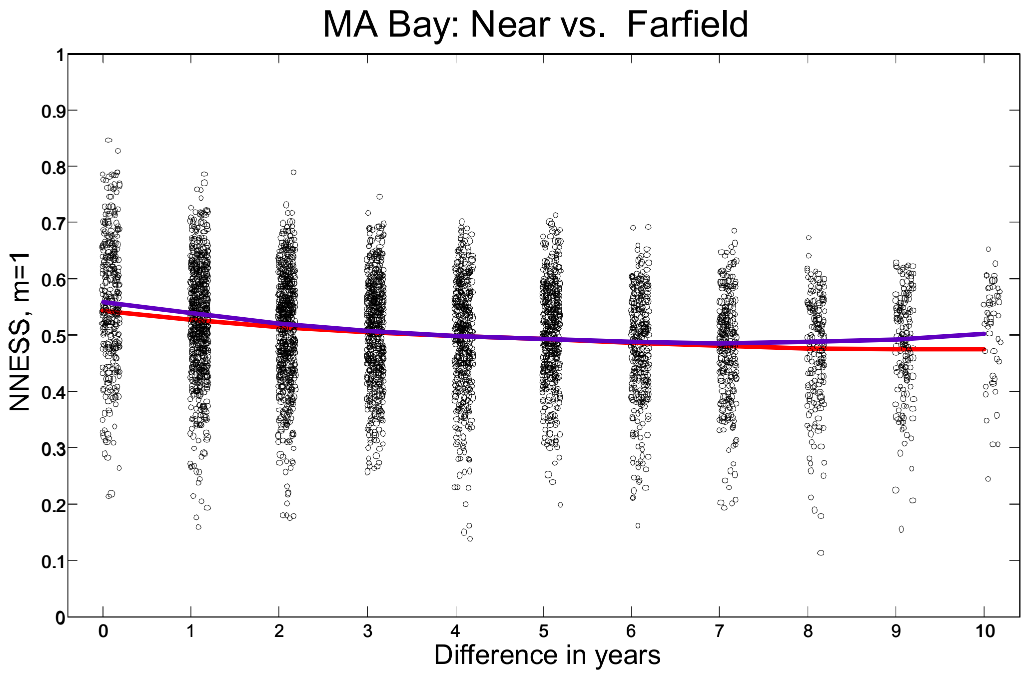
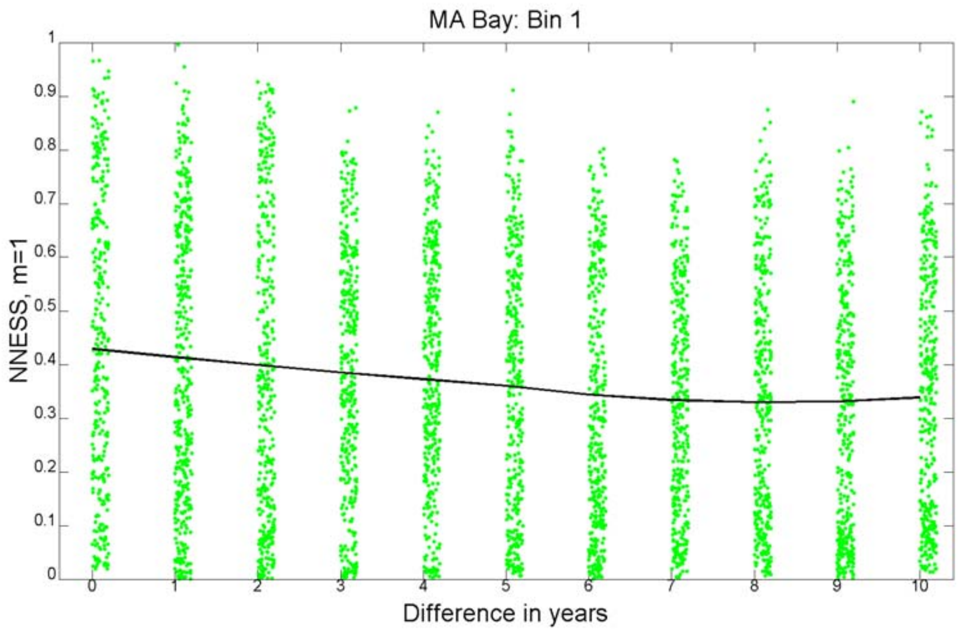


Figure A-5 Time-sorted NNESS biplots (m = 1) and best-fit LOESS lines from Figure A-2 superimposed for nearfield (red line) and farfield (blue line).



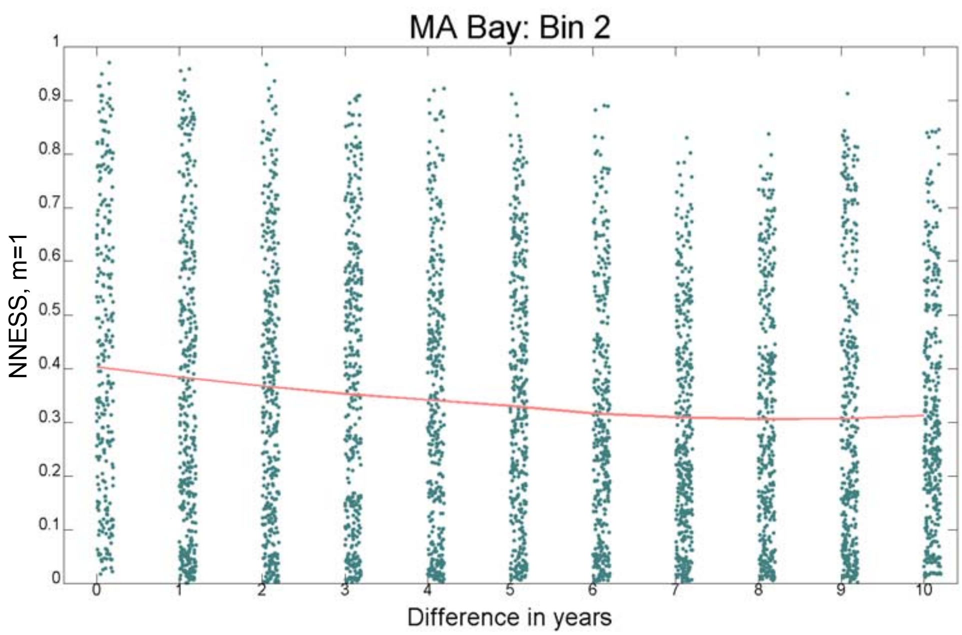


Figure A-6 Time-sorted NNESS values (m = 1) for combined nearfield and farfield data 1992-2002 for the 2 proposed station bins, with best-fit LOESS lines.

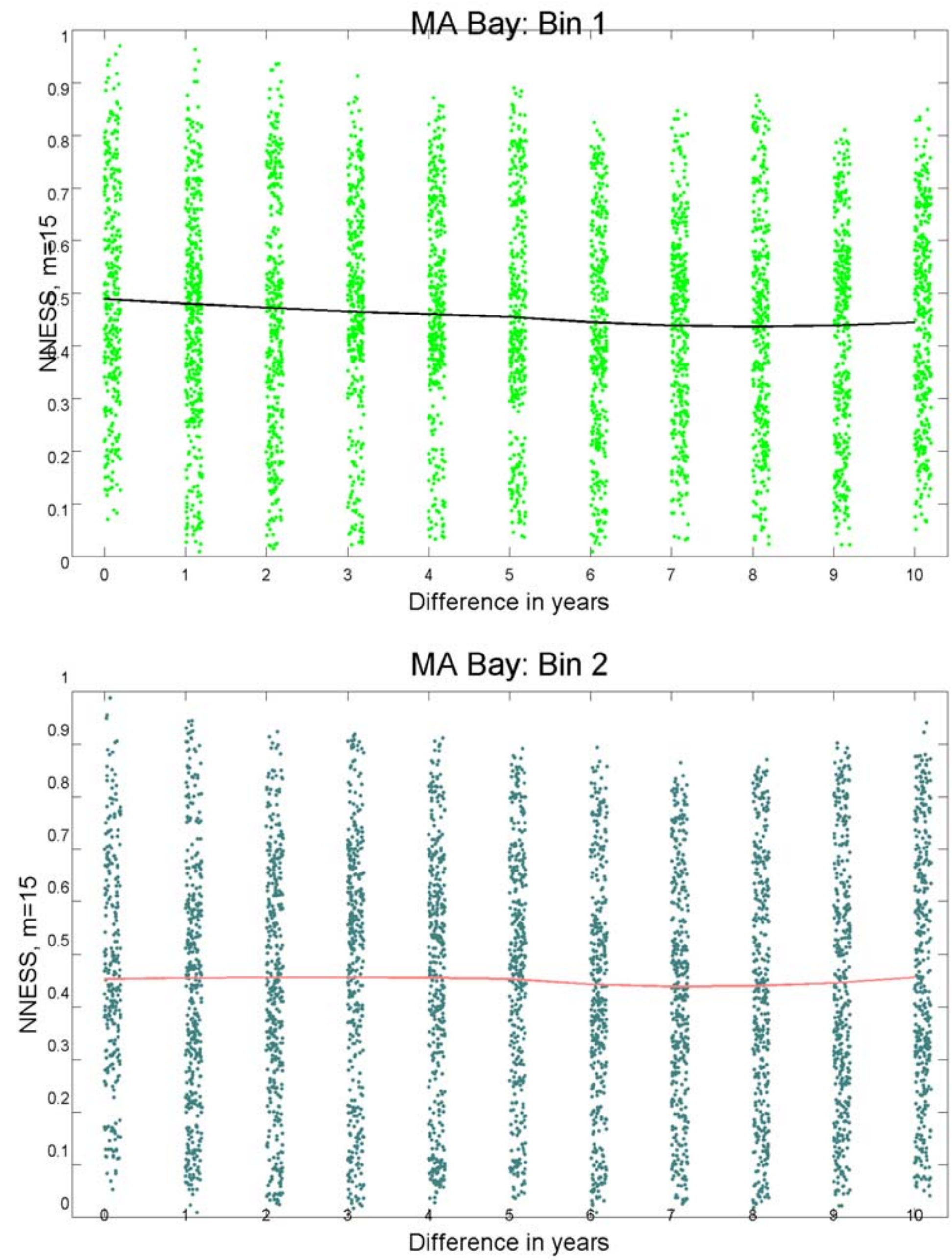
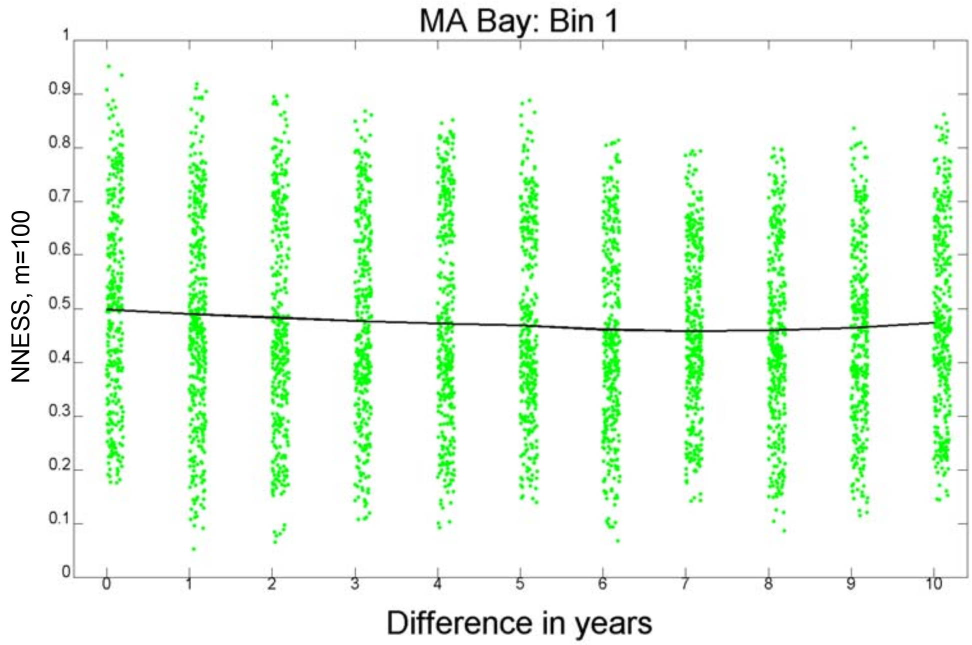


Figure A-7 Time-sorted NNESS values (m = 15) for combined nearfield and farfield data 1992-2002 for the 2 proposed station bins, with best-fit LOESS lines.



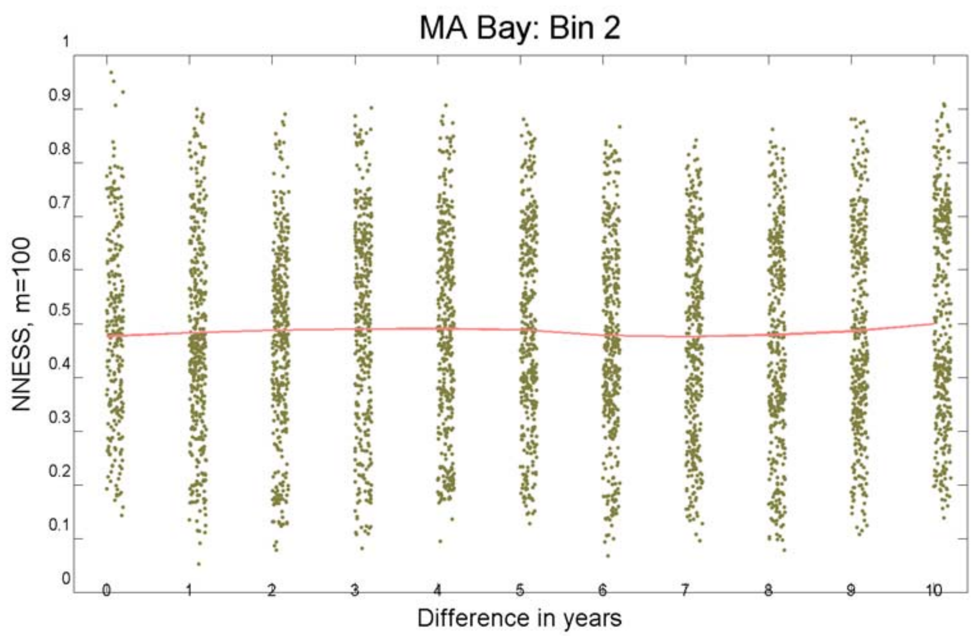


Figure A-8 Time-sorted NNESS values (m = 100) for combined nearfield and farfield data 1992-2002 for the 2 proposed station bins, with best-fit LOESS lines.

A.5 References

Cleveland, W. S. (1993), Visualizing Data. Summit, New Jersey: Hobart Press.

Magurran, A.E. 1988. Ecological diversity and its measurement. Princeton University Press. Princeton.

Smith, W and J.F. Grassle. 1977. Sampling properties of a family of similarity measures. Biometrics 33:283-292.

Trueblood, DD, Gallagher ED Gould DM. 1994. The three stages of seasonal succession on the Savin Hill Cove Mudflat, Boston Harbor. Limnology and Oceanography 39:1440-1454.