ASSESSMENT OF MACROBENTHOS RESPONSE TO SEDIMENT CONTAMINATION IN THE SAN FRANCISCO ESTUARY, CALIFORNIA, USA

BRUCE THOMPSON* and SARAH LOWE
San Francisco Estuary Institute, 7770 Pardee Lane, Oakland, California 94621, USA

(Received 30 April 2003; Accepted 18 February 2004)

Abstract—A multimetric benthic assessment method was developed for two benthic assemblages in the San Francisco Estuary (USA) using data from several monitoring programs collected over five years. Assessment indicators used were total number of taxa, total abundances, oligochaete abundances, number of molluscan taxa, number of amphipod taxa, and Capitella capitata and Streblospio benedicti abundances. Exceedances of the maximum or minimum indicator values in reference samples were used to assess test samples using a weight-of-evidence to obtain an assessment value. Only 2.5% of the samples from the deeper, offshore sites had benthic impacts, 14.3% of the samples from near wastewater discharges had impacts, and 78.3% of the samples from the estuary margins and channels were impacted. Impacted samples from both assemblages had significantly higher mean effects range-median quotient values (mERMq) than reference samples, total organic carbon (TOC) was significantly higher in the impacted samples from the mesohaline assemblage, and percent fines was significantly higher in the impacted samples from the polyhaline assemblage, reflecting the close associations of contaminants with fine sediments and organic material. In samples with mERMq below 0.050, there were no benthic impacts. The incidence of impacts remained low (9.4%) at mERMq below 0.146, but when mERMq was above 0.146, 68.2% of the samples had benthic impacts, and samples with mERMq above 0.740 were always impacted.

Keywords—Benthos Assessment Contaminant response San Francisco Estuary

INTRODUCTION

Benthic organisms are the most common targets in biological assessments of environmental quality [1] because they are an important ecosystem component that provide a primary food source for many fish, birds, and mammals, and affect sediment stability and geochemistry. They also possess attributes appropriate for their use as cost-effective indicators of environmental change, such as limited motility and deposit feeding.

Adaptations of the well-established index of biotic integrity used extensively in fresh water [2], have been applied in marine and estuarine areas along the East Coast of the United States [3,4], in California’s estuaries [5,6], and in San Francisco Bay [7]. In these assessments, benthic indicators (e.g., species diversity, abundance of key taxa) were used in a multimetric index to distinguish impacted from reference benthic conditions. Another assessment approach used multivariate analyses of species composition and abundances to describe assemblage patterns and responses to abiotic variables in the Gulf of Mexico and along the East Coast of the United States [8]. The benthic response index, developed for southern California [9,10], and the benthic assessment methods proposed for Puget Sound (USA) [11] combined the multimetric and multivariate approaches described above.

The objective of this study was to develop and demonstrate a method for assessing impacts of sediment contamination on the benthos at selected sites in the San Francisco Estuary. Benthic assessments generally are conducted as one component of broader sediment assessments that also include consideration of sediment contamination, toxicity, and bioaccumulation [12]. In this paper, the term assessments is used to describe a management tool that uses data to provide a weight-of-evidence about environmental conditions. The term contamination is used to describe the presence of mixtures of trace metals and synthetic organic chemicals above background concentrations.

Background

The benthic assessment procedure described below uses information from a previous study that identified three major macrobenthic assemblages in the San Francisco Estuary [13]. Multivariate analyses identified polyhaline, mesohaline, and oligohaline assemblages. The species composition and abundances of each assemblage primarily reflected responses to the estuarine salinity gradient. Each assemblage was composed of one or more subassemblages that mostly reflected differences in sediment type. The species composition and abundances of each subassemblage tended to be temporally consistent, but the spatial distribution of some subassemblages changed in response to increased freshwater inflows. The estuary margin subassemblage (of the mesohaline assemblage) included samples collected from nearshore subembayments and wetland channels that were inhabited by increased proportions of contamination tolerant and opportunistic taxa, and had elevated sediment contamination and total organic carbon (TOC) levels [7,13] compared to the mesohaline assemblage samples from the deeper portions of the estuary.

Conceptual models of benthic response to sediment contamination

Changes in benthic species composition and abundances often co-occur with changes in sediment contaminant concentrations [14]. However, conceptual models of benthic responses to contamination that reflect current understanding of toxicological, physiological, and ecological mechanisms that control benthic responses are poorly developed. The most com-
increases above biological effects thresholds, abundances decline, with no survival at very high contaminant concentrations. Based on laboratory dose-response models of organisms to contamination, different responses by species may be expected in field samples, depending on the sensitivity or tolerance of a species [9,19]. Contamination-sensitive taxa would be expected to decrease in abundance in the presence of relatively low contaminant concentrations (Fig. 1B). Tolerant species may not be present or exist in low abundances in reference locations, and may increase in abundance at moderate to relatively high contaminant concentrations until their toxic thresholds are exceeded. Benthic response to contamination alone may be moderated by interactions with other taxa. For example, at moderate contaminant concentrations sensitive taxa may be excluded by toxic response, providing a competitive release of space and food to more tolerant taxa, allowing them to increase.

Several other assumptions exist for the models used: Benthic responses to contamination and organic material may be similar because most opportunistic taxa also are contamination tolerant. Responses along geographic gradients should be similar to responses observed when samples that include a range of sediment contamination concentrations are pooled. Other sediment factors usually co-vary with sediment contamination and may confound observed benthic responses.

METHODS AND MATERIALS

Sample collection and analysis

Data from four monitoring programs conducted in the San Francisco Estuary were used in this study (Table 1, Fig. 2). The Regional Monitoring Program for Trace Substances (RMP) sampled during the wet period (January–February) and dry period (August–September) between 1994 and 1998. The Bay Area Clean Water Association’s Local Effects Monitoring Program (LEMP) sampled near the wastewater discharges of the East Bay Municipal Utility District (EBMUD) and the City and County of San Francisco (CCSF) on the same schedule as the RMP from 1994 through 1997. The California Bay Protection and Toxic Clean-Up Program (BPTCP) collected benthic samples at four sites along a suspected contamination gradient in Castro Cove in May 1992, at three prospective reference sites in San Pablo Bay in September 1994, and at 16 suspected toxic hot-spots in April and December, 1997. The California Department of Water Resources (DWR) conducted monthly benthic sampling in the Delta, Suisun, and San Pablo Bays as part of their Compliance Monitoring Program from 1994 to 1997. However, only four of the samples had matching toxicity and sediment contamination data.

A 0.05 m² Pora...
USA) was used to collect the RMP, LEMP, DWR, and 1994 BPTCP samples. Single benthic samples were collected at most sites. Candidate indicator variables were averaged across replicates in samples collected at three RMP sites (1994) and at the DWR sites. The 1992, BPTCP samples were collected using a 0.018 m² grab sampler, and the 1997 BPTCP samples included three sites in adjacent salt marsh channels; the Castro Cove samples included four sites: Point Pinole, CC2, CC4, and EVS4.

At the four DWR sites used, sediment contaminant data from adjacent RMP sites sampled within two weeks was used. Sediment toxicity tests were conducted at many, but not all, of the RMP and BPTCP sites. Ten-day exposures of the amphipod Eohaustorius estuarius to bulk sediments and 48-h bivalve (Mytilus or Crassostrea) embryo exposure to sediment elutriates were conducted [7,21]. Methods used for sediment contamination and toxicity sample collection and analysis are detailed elsewhere [7,21–23].

The effects range-median (ERM) sediment quality guidelines that frequently were associated with biological effects [24] were used to calculate a composite measure of sediment contamination, the mean effects range-median quotient [25]. Depending on the data available, concentrations between 16 and 24 contaminants for which ERM values exist were used. These included eight trace metals (Ag, As, Cd, Cr, Cu, Hg, Pb, Zn), 13 polycyclic aromatic hydrocarbon compounds (acenaphthene, acenaphthylene, anthracene, benzo[a]pyrene, benzo[a] anthracene, chrysene, dibenz[a]anthracene, fluoranthene, fluorene, 2-methylnaphthalene, naphthalene, phenanthrene, pyrene), p,p'-dichlorodiphenyldichloroethylene (DDE), total DDTs, and total polychlorinated biphenyls. Mean ERM quotients computed using these components have been used in previous studies of benthic impacts [26,27].

Data analyses were conducted using the statistical analysis system [28]. Correlation and multiple regression analyses were used to evaluate relationships between benthic indicators and abiotic variables, particularly the proportions of variance in indicator values accounted for by sediment contamination (mERMq) when salinity, percent fines, and TOC also were included in the analysis. Because the expected responses of the indicators to the abiotic variables were curvilinear, the data were transformed prior to analysis (natural log or arcsine). The regression model that included the combination of transformed and/or untransformed abiotic variables (independent variables) that provided the highest $R^2$ value for each assessment indicator (dependent variable) was used. These analyses were used only to evaluate the relative contributions of selected abiotic variables to indicator variation, not for predictions of indicator responses. The Wilcoxon two-sample test with ranked data was used to compare samples statistically.

Benthic assessment procedure

The assessment procedure used a combination of the two general approaches described in the Introduction section. In a previous paper, multivariate analyses assigned samples to benthic assemblages [13]. The identification of assemblages is an important step in bioassessments. Because habitat conditions and species composition within assemblages are relatively homogeneous, the development of assessments for each assemblage minimizes the variability in benthic responses to large differences in salinity or sediment type found in different assemblages. The assessment method was developed for two major benthic assemblages: The polyhaline and mesohaline assemblages (including the estuary margin subassemblages) because they had adequate sample sizes and accompanying sediment toxicity and contamination data. A multimetric set of benthic indicators was used for the assessments as described below (see Results section).

RESULTS

Identification of candidate benthic indicators

Candidate indicators were identified from the literature and from extensive testing and evaluation of the data. Many of the
candidate indicators have been used in benthic assessments of other estuaries (Table 2). Indicators were identified for possible use in two steps of the assessments procedure: Screening to identify reference samples and benthic assessment indicators. The number of taxa and total abundance per sample are the most commonly used indicators of benthic assemblage diversity and structure [2,14,18]. Most amphipods are sensitive to contamination [29]. However, *Grandidierella japonica* inhabited sediments with very high DDT concentrations at the Richmond Harbor Superfund site [30]. Thus, it was considered unsuitable for use as an indicator. *Amphela abita* is often the dominant organism in benthic samples in the San Francisco Estuary [13], and commonly is used in sediment bioassays [31]. However, the life history of this species produces highly variable seasonal abundances in the San Francisco Estuary [32], and it was not considered suitable as an assessment indicator in this study. The number of amphipod taxa in the polyhaline assemblage was selected as a candidate indicator. Mollusks commonly were collected in the estuary, are sensitive to contaminants [33], and were absent from some of the contaminated BPTCP sites, suggesting possible benthic impacts that were not distinguished by the classification analysis. No locations are free of sediment contamination in the estuary [39], and sediment tox-
The presence of amphipods [29], oligochaetes contributed less than 30% of total abundances, and pollution-tolerant taxa contributed less than 50% of the total taxa. Because no reference threshold values for the latter two indicators were found in the literature, the thresholds used were applied based on experience, professional judgment, and examination of the data. Samples that did not meet at least two of the criteria were considered to exhibit possible contaminant impacts and were eliminated as reference samples. Exceeding only one criterion provided a benefit-of-doubt to samples that may have exceeded one criterion for reasons unrelated to contamination. For example, the absence of amphipods at Davis Point (BD41) probably was due to the sandy conditions there.

Twenty of the polyhaline assemblage samples from six sites collected in both wet and dry seasons, 1994 to 1998, were identified as reference samples. All of the samples eliminated were due to sediment toxicity. The reference samples had less than 28% oligochaetes, but nine of the samples had relative proportions of tolerant taxa above 0.5 (up to 0.63), the only reference-screening criterion exceeded. Nine of the candidate assessment indicators were not correlated significantly with mERMq. Sediment contamination (mERMq) accounted for significant portions (22–88%) of the variation for most of the polyhaline assessment indicators except oligochaetes and C. capitata (Table 3A). Sediment contamination (mERMq) contributed most to indicator variation for number of taxa, total abundance, and number of amphipod taxa, but salinity accounted for more of the variation for oligochaete and C. capitata abundances. In the mesohaline assemblage, mERMq accounted for 15 to 95% of indicator variation. Total organic carbon accounted for more of the variation in indicator responses than mERMq, except for oligochaete abundances. This was due to the elevated TOC values (mean = 2.15%) in the estuary margins subassemblage where mERMq values also were the highest (mean = 0.297); TOC and mERMq were correlated significantly in those samples (Spearman’s rho = 0.631; p = 0.005).

These evaluations showed that either all of the candidate indicators except C. capitata were significantly correlated with sediment contamination or sediment contamination was a significant component of the regression models. In the mesohaline assemblage, TOC appeared to have a greater influence on the indicators than sediment contamination, and these results will be considered in the interpretation of assessment results.

Selection of assessment indicators

Based on the results of the evaluations above and previous use in benthic assessments in other regions of the United States, four indicators were selected for use in benthic assessments in the polyhaline assemblage and five were selected for use in the mesohaline assemblages (Table 2). Consistent with the principles of index of biotic integrity-type assessments [2], indicators of the structure of the assemblage, higher taxa, and indicator species were selected.

Oligochaete abundances in the polyhaline assemblage were not selected because they did not exhibit a strong response to sediment contamination. Additionally, their range of values in the reference samples was between 0 and 182, such that none of the test samples exceeded that range, making them unsuitable as an assessment indicator. Abundances of C. capitata
were selected for use in the assessments. Although the data used showed only weak responses to sediment contamination, their value as an indicator of impacted benthos in other areas is well documented.

The selection of indicators for total abundance and abundance of selected indicator taxa may be somewhat redundant, especially if the indicator taxon contributes a large portion of total abundance. As will be shown, when used together in benthic assessments of samples from each assemblage, these indicators distinguished samples with benthic impacts from reference samples.

**Calculation of indicator reference ranges**

The range of values (minimum, maximum) for each benthic assessment indicator was calculated using the reference samples from each assemblage (Table 4). These maximum and minimum reference values include variation in space (among assemblage samples) and over time (1994–1998). The reference ranges will be used to assess test benthic samples by comparing the indicator values in the test samples to the reference ranges to determine whether the test sample values are within the reference range, or outside of it. For example, a sample collected from the polyhaline assemblage with fewer than 16 taxa would be below the reference range providing one component for a weight-of-evidence of a possible impact. The upper range value also is useful. The conceptual models showed that samples with moderate contamination might have higher numbers of taxa than the reference samples (Fig. 1A). Thus, if a polyhaline assemblage sample had more than 66 taxa, it could indicate a moderate impact.

The three contamination-tolerant indicator taxa (oligochaetes, *C. capitata, S. benedicti*) had 0 as the lower range value. These indicator taxa often were not present in reference samples and were expected to increase with moderate to high contamination. Exceeding the upper reference range would show expected responses to contamination.

**Assessments of benthic samples**

Benthic assessments were conducted on 122 samples from the polyhaline and mesohaline assemblages. Sample indicator values that exceeded a reference range were considered a hit, and the sum of the hits from comparisons of all indicators assessed in each sample produced a weight-of-evidence for the degree of benthic impact, the assessment value (AV). Samples with none or one of the indicators outside the reference range were considered to be unimpacted. As in screening reference samples, exceeding only one reference range provided a benefit-of-doubt to sites that may exceed a reference range for reasons unrelated to contamination. Samples with two indicators outside their reference ranges (AV = 2) were considered to be slightly impacted, samples with AV = 3 were considered to be moderately impacted, and samples with AV = 4 or 5 were considered to be severely impacted.

**Polyhaline assemblage.** Only one of 26 RMP polyhaline assemblage samples was assessed as slightly impacted (AV = 2; Table 5A). Six of the 39 LEMP sites near two wastewater outfalls were assessed as impacted. Four samples from near the CCSF discharge were slightly to moderately impacted (AV = 2,3). All four samples were from the wet sampling periods in 1995 or 1996 and had numbers of taxa, total abundances,
Table 5. Assessment results from polyhaline assemblage and mesohaline assemblage samples, showing impacted samples. The \( mER\text{Mq} \) = mean effects range-median quotient; \( AV \) = assessment value. \( A \) = value was above the reference range; \( B \) = value was below the reference range.

### Polyhaline assemblage

<table>
<thead>
<tr>
<th>Polyhaline assemblage</th>
<th>Date</th>
<th>No. taxa</th>
<th>Total abundance</th>
<th>Amphipod taxa</th>
<th>Capitella capitata</th>
<th>AV</th>
<th>mER\text{Mq}</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB15</td>
<td>2/15/1994</td>
<td>24</td>
<td>81 B</td>
<td>0 B</td>
<td>1</td>
<td>2</td>
<td>0.0766</td>
</tr>
<tr>
<td>CCSF04</td>
<td>2/13/1995</td>
<td>13 B</td>
<td>27 B</td>
<td>1 B</td>
<td>0</td>
<td>3</td>
<td>0.1977</td>
</tr>
<tr>
<td>CCSF04</td>
<td>3/26/1996</td>
<td>25</td>
<td>68 B</td>
<td>1 B</td>
<td>0</td>
<td>2</td>
<td>0.1092</td>
</tr>
<tr>
<td>CCSF05</td>
<td>3/26/1996</td>
<td>15 B</td>
<td>37 B</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0.1595</td>
</tr>
<tr>
<td>CCSF06</td>
<td>3/26/1996</td>
<td>21</td>
<td>88 B</td>
<td>1 B</td>
<td>14 A</td>
<td>3</td>
<td>0.1922</td>
</tr>
<tr>
<td>EBMUD04</td>
<td>9/23/1994</td>
<td>60</td>
<td>4,866 A</td>
<td>11</td>
<td>16 A</td>
<td>2</td>
<td>0.0655</td>
</tr>
<tr>
<td>EBMUD04</td>
<td>8/16/1995</td>
<td>49</td>
<td>16,760 A</td>
<td>9</td>
<td>26 A</td>
<td>2</td>
<td>0.0830</td>
</tr>
<tr>
<td>EBMUD05</td>
<td>8/16/1995</td>
<td>43</td>
<td>18,723 A</td>
<td>10</td>
<td>18 A</td>
<td>2</td>
<td>0.1176</td>
</tr>
<tr>
<td>EBMUD06</td>
<td>8/16/1995</td>
<td>40</td>
<td>14,101 A</td>
<td>9</td>
<td>37 A</td>
<td>2</td>
<td>0.0878</td>
</tr>
<tr>
<td>EBMUD05</td>
<td>8/19/1997</td>
<td>58</td>
<td>3,948 A</td>
<td>12 A</td>
<td>6</td>
<td>2</td>
<td>0.0941</td>
</tr>
<tr>
<td>ZM-1</td>
<td>12/2/1997</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1.7686</td>
</tr>
<tr>
<td>ZM-2</td>
<td>12/3/1997</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0.5614</td>
</tr>
</tbody>
</table>

### Mesohaline assemblage

<table>
<thead>
<tr>
<th>Mesohaline assemblage</th>
<th>Date</th>
<th>No. taxa</th>
<th>Total abundance</th>
<th>Molluscan taxa</th>
<th>Oligochaete abundance</th>
<th>Streblus benedicti</th>
<th>AV</th>
<th>mER\text{Mq}</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA21</td>
<td>2/16/1994</td>
<td>13</td>
<td>1,603 A</td>
<td>2</td>
<td>61 A</td>
<td>0</td>
<td>2</td>
<td>0.1614</td>
</tr>
<tr>
<td>WBCC 2A</td>
<td>2/28/1995</td>
<td>11</td>
<td>782</td>
<td>1</td>
<td>596 A</td>
<td>76 A</td>
<td>2</td>
<td>0.0826</td>
</tr>
<tr>
<td>WBCC 2B</td>
<td>3/1/1995</td>
<td>21 A</td>
<td>1,793 A</td>
<td>2</td>
<td>965 A</td>
<td>209 A</td>
<td>4</td>
<td>0.0811</td>
</tr>
<tr>
<td>WBC 3B</td>
<td>3/1/1995</td>
<td>15</td>
<td>1,976 A</td>
<td>2</td>
<td>1,853 A</td>
<td>298 A</td>
<td>3</td>
<td>0.0756</td>
</tr>
<tr>
<td>CC4</td>
<td>5/1/1992</td>
<td>22 A</td>
<td>1,291 A</td>
<td>5 A</td>
<td>11</td>
<td>25</td>
<td>3</td>
<td>0.0991</td>
</tr>
<tr>
<td>P8pinole</td>
<td>5/1/1992</td>
<td>19 A</td>
<td>1,302 A</td>
<td>3</td>
<td>22</td>
<td>34</td>
<td>2</td>
<td>0.0509</td>
</tr>
<tr>
<td>CCR4</td>
<td>12/3/1997</td>
<td>33 A</td>
<td>3,714 A</td>
<td>3</td>
<td>481 A</td>
<td>551 A</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>PSEnd</td>
<td>4/2/1997</td>
<td>9</td>
<td>194</td>
<td>0 B</td>
<td>87 A</td>
<td>11</td>
<td>2</td>
<td>0.2096</td>
</tr>
<tr>
<td>MC-1</td>
<td>4/17/1997</td>
<td>15</td>
<td>5,321 A</td>
<td>0 B</td>
<td>4,963 A</td>
<td>0</td>
<td>3</td>
<td>1.3930</td>
</tr>
<tr>
<td>MC-End</td>
<td>4/1/1997</td>
<td>27 A</td>
<td>656</td>
<td>6 A</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0.1562</td>
</tr>
<tr>
<td>MC-Mid</td>
<td>4/1/1997</td>
<td>9</td>
<td>141</td>
<td>0 B</td>
<td>3</td>
<td>61 A</td>
<td>2</td>
<td>0.4442</td>
</tr>
<tr>
<td>IC-End</td>
<td>4/1/1997</td>
<td>20 A</td>
<td>452</td>
<td>0 B</td>
<td>54 A</td>
<td>126 A</td>
<td>4</td>
<td>0.3224</td>
</tr>
<tr>
<td>SLB-3</td>
<td>4/17/1997</td>
<td>40 A</td>
<td>5,014 A</td>
<td>10 A</td>
<td>541 A</td>
<td>1,208 A</td>
<td>5</td>
<td>0.3264</td>
</tr>
<tr>
<td>SLB-4</td>
<td>4/17/1997</td>
<td>34 A</td>
<td>3,489 A</td>
<td>9 A</td>
<td>165 A</td>
<td>708 A</td>
<td>5</td>
<td>0.7422</td>
</tr>
<tr>
<td>SLB-5</td>
<td>4/17/1997</td>
<td>27 A</td>
<td>4,847 A</td>
<td>6 A</td>
<td>205 A</td>
<td>710 A</td>
<td>5</td>
<td>0.2014</td>
</tr>
<tr>
<td>SLB-6</td>
<td>4/17/1997</td>
<td>37 A</td>
<td>4,561 A</td>
<td>7 A</td>
<td>98 A</td>
<td>154 A</td>
<td>5</td>
<td>0.1468</td>
</tr>
<tr>
<td>SLB-7</td>
<td>4/16/1997</td>
<td>30 A</td>
<td>3,845 A</td>
<td>7 A</td>
<td>394 A</td>
<td>490 A</td>
<td>5</td>
<td>0.6111</td>
</tr>
</tbody>
</table>

and amphipod abundances outside reference ranges; one sample had *C. capitata* abundances above the reference range. Five samples from near the EBMUD discharge were slightly to moderately impacted. However, three of those samples had very high densities of the amphipod *Monocorophium ascherusicum* (up to 246,880 m\(^{-2}\)), such that amphipod abundance and total abundance reference ranges were exceeded. The genus *Monocorophium* is considered to be sensitive to sediment contamination [15,30], and the episodic influx of *M. ascherusicum* (which also occurred at RMP site BC11, assessed as unimpacted) probably was not related to sediment contamination. Therefore, those samples were not considered to be impacted. The two slightly impacted EBMUD samples had total abundances and amphipod or *C. capitata* abundances outside the reference ranges and all were from the dry sampling period. Sediment contamination in the LEMP samples was significantly higher (average \( mER\text{Mq} = 0.121 \)) than in the RMP polyhaline assemblage samples (average \( mER\text{Mq} = 0.084 \); Wilcoxon \( p < 0.001 \)), but TOC was not significantly different between the two sets of samples \((p = 0.691)\). Thus, the significantly higher \( mER\text{Mq} \) values at the LEMP sites were accompanied by a higher incidence of benthic impact (15.4%) than at the RMP polyhaline sites (3.8%). Two BPTCP samples in Zeneca Marsh were devoid of benthic organisms (Table 5A). Although they did not exceed the reference range for *C. capitata* abundances (0 lower range), the severe impact was obvious.

*Mesohaline assemblage.* Only one of 23 RMP mesohaline assemblage samples was classified as slightly impacted \((AV = 2; Table 5B)\). Three of the four RMP samples from the China Camp tidal marsh channels (Station WBCC, estuary margin subsample) were impacted to varying degrees with oligochaete and *S. benedicti* abundances above reference ranges. Although the tidal marsh samples had slightly higher \( mER\text{Mq} \) values \((mean = 0.079)\) than the mesohaline reference samples \((mean = 0.069)\), they were not significantly different \((Wilcoxon \( p = 1.0 \)) , but TOC in the marsh samples \((mean = 2.55)\) was significantly higher than in the mesohaline reference samples \((mean = 0.91; p < 0.019\). Three LEMP samples from near the CCSR outfall in the Central Bay following flood flows in January 1997 were classified in the mesohaline assemblage, and were unimpacted, and all four of the DWR samples assessed were unimpacted. Four BPTCP samples from an abandoned refinery outfall gradient into Castro Cove sampled in 1992 were assessed. However, the two samples inside the Cove, closest to the old outfall, were not impacted \((CC2, EVS4)\). Despite elevated sediment contamination at site EVS4 \((mER\text{Mq} = 0.635)\) nearest to the old outfall, all assessment indicators were within reference ranges. Another 14 BPTCP samples collected in April 1997 were assessed. Eleven of those samples showed benthic impacts. One sample from a potential BPTCP reference site at Carlson Creek was severely impacted \((AV = 4)\), and the samples from a three-site gradient at Mission Creek in San Francisco were slightly to moderately impacted \((AV = 2.3)\). Two samples from the channels at Islais Creek and Peyton Slough were impacted \((AV = 2.4, respectively)\). Five samples from San Leandro Bay (SLB) were severely
impacted (AV = 5); all assessment indicators were above the mesohaline assemblage reference ranges. Both mERMq and TOC in the SLB samples (mERMq mean = 0.406; TOC mean = 3.27) were significantly higher than in the mesohaline reference samples (mERMq mean = 0.069, TOC mean = 0.912; Wilcoxon p < 0.015).

**Evaluation of the assessment procedure**

Samples with benthic impacts (AV = 2–5) had significantly higher mERMq levels than the reference samples in each assemblage (Table 6). Total organic carbon in sediments was not significantly different between impacted and reference samples in the polyhaline assemblage, but TOC was significantly higher in the impacted samples than in the reference samples in the mesohaline assemblages. Percent fine sediments were significantly higher in the impacted samples than in the reference samples in the polyhaline assemblage, but not in the mesohaline assemblage samples. Salinities were similar in reference and impacted samples in both assemblages. Because toxicity was a reference sample screening criterion, the reference samples had no sediment toxicity compared to 100% and 85.7% toxicity in the impacted samples in the mesohaline and polyhaline assemblages, respectively, corresponding with the sediment contamination patterns and with the benthic assessment results (Table 6).

The incidence of benthic impacts in the estuary corresponded with increasing mERMq values. No benthic impacts were in samples (n = 13) with mERMq values below 0.051. Impacts occurred in 9.4% of the samples (n = 106) with mERMq values between 0.51 and 0.146, in 63.2% of the samples (n = 19) with mERMq values between 0.147 and 0.635, and in all samples (n = 3) with mERMq values above 0.742. The highest mERMq value from the reference samples was 0.146, corresponding to the step increase in incidence described above. Only 8.4% of the samples below 0.146 were impacted, but 68.2% of samples above that value were impacted. Therefore, a mERMq value of 0.146 appears to provide a reasonable guideline for reference sediment contamination levels in the San Francisco Estuary, below which samples were not toxic (a reference sample screening criterion) and there was a low risk of benthic impacts.

Many of the impacted samples had at least one assessment indicator value above the reference range in both assemblages. Consistent with the conceptual models, moderate impacts may result in elevated numbers of taxa or abundances. However, the five severely impacted (AV = 5) San Leandro Bay samples had all five indicators above reference ranges.

**DISCUSSION AND CONCLUSION**

The benthic assessment procedure used a suite of benthic indicators to clearly distinguish impacted samples from reference samples in the two largest assemblages in the San Francisco Estuary. The results were consistent with current conceptual models of benthic impact, reflected increasing sediment contaminant concentrations, and were consistent with matching sediment toxicity results.

The assessments showed that the highest incidence (78%) and most severe benthic impacts occurred at sites in the subembayments, coves, and channels along the margins of the estuary. Samples from the deeper areas, farther offshore, in both the polyhaline and mesohaline assemblages, had a much lower incidence (2.4%) of slight impacts. Benthic impacts occurred at 14.3% of the samples from near the wastewater discharges. However, because none of the sites used in this study were selected randomly, these patterns of impact may simply reflect the goals of each program from which data were used. The RMP and DWR sites were selected as background or representative sites, and the BPTCP and LEMP samples were selected to evaluate toxic hot spots and wastewater discharge conditions, respectively. Thus, although the higher incidence of benthic impacts in the estuary margins and LEMP samples strongly suggests that these areas are much more impacted than areas farther offshore, further random sampling will be required to confirm the spatial patterns of impacts.

In the mesohaline assemblage, TOC appeared to have greater influence on most assessment indicators (except oligochaetes), than mERMq (Table 4). However, both mERMq and TOC were significantly higher in the impacted mesohaline samples than reference samples (Table 6). The estuary margins receive direct inputs from numerous tributaries and storm drains that deliver sediment, organic material, and contaminants [42]. Aquatic vegetation growing along the estuary margins also provides detrital organic material inputs. Elevated sediment contamination and TOC co-occurred with increased abundances of opportunistic benthic organisms, most of which also are contaminant tolerant (e.g., oligochaetes, *S. benedicti*). Due to these elevated levels and close association between sediment contamination and TOC along the estuary margins, the apparent benthic impacts in those samples probably were due to a combination of influences from elevated contamin-
nation and TOC. In the polyhaline assemblage, mERMq had more influence on each indicator than percent fines, but the impacted samples had significantly higher values of both of those variables than reference samples. Therefore, impacts in that assemblage probably were due to the combined influence of fine sediments and accompanying increased contamination.

The results of the assessments of the BPTCP samples conducted in this study identified more impacted samples than were identified in the BPTCP report using a different benthic assessment method [7]. They identified five impacted samples: Two from Zeneca [Stege] Marsh, two from Islais Creek (IC), and one from Mission Creek. The assessments conducted in this report also showed impacts in those samples, but additionally showed impacts at two other Mission Creek sites, all five San Leandro Bay sites, Carlson Creek, and Peyton Slough.

The results of this study showed benthic impacts at slightly higher mean ERM quotient values than in Atlantic and Gulf coast estuaries where the mean ERM quotient threshold for moderate or medium benthic impact risks (31–52% of samples) ranged between 0.013 and 0.022, and for high-impact risks (55–85% of samples) ranged between 0.036 and 0.098 [43]. Mean ERM quotient values above 0.20 resulted in marked decreases in number of benthic taxa and arthropod abundance in Florida’s Biscayne Bay (USA) [26]. In comparison, mERMq values related to sediment toxicity usually are much higher than for benthic impacts; sediment samples were toxic (76.5%) when ERM quotients were above 1.0 [25], presumably due to greater sensitivity of benthic community indicators to sediment contamination, than standard toxicity test species. However, amphipod toxicity test results often corresponded to significant alterations to benthic indicators [41].

Although it was not possible to identify sediment contamination as the sole factor influencing benthic impacts, the assessment results reflected increasing sediment contamination, which appeared to have a major influence on the benthos. In most estuarine systems, benthic assessments may not distinguish contaminant effects from effects due to other sediment factors, such as TOC due to the close association between contamination and organic material. Therefore, management response to the results of this, or any benthic assessment, should be to conduct more detailed investigations at sites where benthos appear to be impacted to ascertain that contaminants are persistent and a factor in the apparent impact. Consistent with the recommended usage of the Sediment Quality Triad, each component of the triad (sediment contamination, toxicity, and benthos) contributes evidence of sediment conditions [12]. The benthic component contributes an evaluation of an important ecosystem component and should be interpreted in conjunction with the other components to provide a comprehensive assessment of the condition of sediments in the estuary.

Acknowledgement—We thank the Bay Area Clean Water Association (BACWA), R. Fairey, and M. Vayssières for use of their data. BACWA contributed financial support for portions of this study. Taxonomy was conducted by A. Navarette, M. Kellogg, K. Langan-Cranford, P. McGregor, B. Sak, W. Fields, J. Oaken, and P. Slattery. The benthic samples were sorted by S. McCormick, P. Salop, and D. Bell. M. Kellogg and Captain G. Smith on the research vessel and D. Johnston coordinated the collection of the RMP samples. Sediment analyses were conducted by R. Flegal, G. Scelfo, S. Hibdon, L. Butler, C. Davis, J. Sericano, and F. Rodigari. This is San Francisco Estuary Regional Monitoring Program Contribution 60.

REFERENCES


