Levels of agreement among experts using Best Professional Judgment to assess mesohaline and tidal freshwater benthic macrofaunal condition in the San Francisco Estuary and Delta

Bruce Thompson[1], Stephen B. Weisberg, Aroon Melwani[1], Sarah Lowe[1], J. Ananda Ranasinghe, Donald B. Cadien[2], Daniel M. Dauer[3], Robert J. Diaz[4], Wayne Fields[5], Michael Kellogg[6], David E. Montagne[7], Peter R. Ode[8], Donald J. Reish[9] and Peter N. Slattery[10]

Abstract

Benthic indices to support aquatic environmental condition assessments have been more effectively developed for higher than lower salinity habitats. Here we quantify agreement among benthic experts using best professional judgment to assess community condition of mesohaline and tidal freshwater samples from the San Francisco Estuary and Delta, and compare that to a previous study for San Francisco Estuary polyhaline samples. Benthic species abundance data from 20 sites in each habitat were provided to 7 tidal freshwater and 8 mesohaline experts who ranked the samples from best to worst condition and placed the samples into 4 condition categories. The average correlation among expert's condition rankings was only 0.38 and 0.29 in the mesohaline and tidal freshwater habitats, respectively, compared to 0.92 in the previous polyhaline study. Pair-wise agreement among expert condition categories averaged 41 and 39%, compared to 70% in the polyhaline. Based on post-exercise discussions among the experts, the differences in agreement among habitats appears related to the use of different indicator taxa and to disturbance regimes in the lower salinity habitats that select for higher proportions of tolerant taxa, confounding assessments at the current level of understanding of benthic response in these habitats. Regardless of the reason, the absence of a clear conceptual model and agreement among benthic ecologists about benthic condition makes index development more difficult in lower salinity habitats.

Introduction

Benthic community condition is widely used in aquatic systems to assess the effects of numerous stressors, including physical disturbance, organic loading, and chemical contamination on the biota (Dauer *et al*. 2000, Borja *et al*. 2003, Diaz *et al*. 2004, Muxika *et al*. 2005, Borja and Dauer 2008, Pinto *et al*. 2009). Benthic macrofauna are commonly used because they are sensitive and relatively immobile residents in sediments. Their use in assessments has expanded considerably over the last decade as benthic indices have become more preva-

[1] *San Francisco Estuary Institute, Oakland, CA*
[2] *Sanitation Districts of Los Angeles County, Whittier, CA*
[3] *Old Dominion University, Department of Biological Sciences, Norfolk, VA*
[4] *Virginia Institute of Marine Science, Gloucester Point, VA*
[5] *Hydrozoology, Newcastle, CA*
[6] *City and County of San Francisco, San Francisco, CA*
[7] *Independent Researcher, Penn Valley, CA*
[8] *California Aquatic Bioassessment Laboratory, Rancho Cordova, CA*
[9] *California State University, Department of Biological Sciences, Long Beach, CA*
[10] *Moss Landing Marine Laboratories, Moss Landing, CA*

lent (Marques *et al*. 2009). Benthic indices summarize the complex species composition information in a sample and provide a numerical scale of community condition from good to bad that facilitates interpretation in a management context.

However, index development has occurred primarily for polyhaline and euhaline environments. While benthic indices are widely used in freshwater streams (USEPA 2002) and in some lakes (Schloesser *et al*. 1995, Hartig *et al*. 1997, USEPA 1998, Blocksom *et al*. 2002), they have yet to be successfully developed for low salinity habitats in estuaries. Weisberg *et al*. (1997) developed separate indices for seven Chesapeake Bay benthic habitats, but found reduced levels of success with decreasing salinity and particularly poor index validation in tidal freshwater. Alden *et al*. (2002) described further development of a Chesapeake Bay tidal freshwater benthic index which met with limited success despite investments in targeted data collection to create an improved index calibration data set. Dauvin (2007) and Dauvin and Ruellet (2009) described some of the potential impediments to developing benthic indices in transitional low salinity habitats.

Benthic indices generally do not represent new ways of thinking about benthic communities, but are mathematical representations based on conceptual models of perturbation effects of the characteristics of impaired and unimpaired communities. The conceptual models are generally based on expert agreement or consensus about these characteristics. Two recent papers (Weisberg *et al*. 2008, Teixeira *et al*. 2010) address the extent to which experts using best professional judgement (BPJ) as a means of determining expectations for index performance agree in their assessment of benthic communities in polyhaline and euhaline environments. However, no such studies have been conducted in lower salinity habitats. The objectives of this study were to determine the levels of expert agreement about benthic condition in the mesohaline (moderate salinity) habitats of the San Francisco Estuary and the tidal freshwater (equivalent to limnetic in the Venice system of terminology) habitats of the Sacramento- San Joaquin River Delta, and compare the levels of agreement among experts to those achieved in a similar study (Weisberg *et al*. 2008) previously conducted in the San Francisco Estuary polyhaline habitat.

Methods

Eight experts were provided species composition, abundance, and basic habitat measures (salinity, total organic carbon, and sediment grain-size) for 20 benthic samples from the mesohaline assemblage of the San Francisco Estuary. Additionally, 7 experts were provided data for 20 tidal freshwater samples from the Sacramento-San Joaquin River Delta. Five of the experts evaluated samples from both assemblages. The experts were selected to represent a range of affiliations and contributed as coauthors of this paper. All of the experts have at least 20 years of experience in interpretation of benthic community data from a wide variety of habitats throughout the United States (US); 3 of the experts have direct experience in the San Francisco Estuary and/or Delta.

The sites were systematically selected to represent the entire range of geography and sediment conditions within each habitat. Sediment variables used in the selection process were sediment contamination (mean Effects Range-Median quotient (mERMq); Long *et al*. 1995), percent fine sediments (<63 μm), and percent total organic carbon. While chemical contamination was used to ensure that a range of site conditions was included in the assessment, the experts were not provided the chemical data.

The experts were asked to rank the relative condition of each site from best to worst within each assemblage, based on data provided. They were also asked to assign each site to one of four categories of benthic condition: 1) undisturbed - a community at a "least disturbed" or "undisturbed" site that may be considered a "reference" condition; 2) low disturbance - a community that shows some indication of disturbance, but could be within measurement error of undisturbed; 3) moderate disturbance - a community that shows evidence of physical, chemical, natural, or anthropogenic disturbance; and 4) highly disturbed - a community with an obvious high level of disturbance. The experts were also asked to list the benthic metrics and indicators they used to determine their rankings and categories, and to rate the importance of those attributes.

Results

Mesohaline Samples

The relative rankings of the samples were poorly correlated among experts, with an average Spearman correlation of 0.38 and only 57% of the expert pairs significantly correlated (Table 1). The rankings of

Table 1. Spearman rank correlation coefficients between expert mesohaline sample rankings. Positive significant (p <0.05) correlations are bold.

	A	B	C	D	E	F	G
A							
B	0.333						
C	**0.639**	-0.187					
D	**0.780**	**0.493**	**0.605**				
E	**0.505**	-0.040	**0.508**	0.392			
F	**0.740**	-0.015	**0.647**	**0.651**	**0.635**		
G	**0.451**	-0.245	**0.478**	0.211	**0.812**	**0.680**	
H	0.032	-0.266	0.309	0.204	**0.496**	0.369	**0.565**

Expert B were inversely correlated with five of the other experts' rankings. All of the experts' ranks except B and H were significantly correlated with the median ranks (Table 2), indicating that that the ranks provided by those experts were much different than the others.

Table 2. Spearman correlation coefficients between expert mesohaline sample rankings and the median rank. Significant (p<0.05) correlations are bold.

A	B	C	D	E	F	G	H
0.734	-0.031	**0.577**	**0.608**	**0.599**	**0.738**	**0.495**	0.317

There were no samples for which all of the experts agreed on the condition category and for only nine of 20 samples did even a majority of the experts agree (Table 3). Most of the samples for which there was majority agreement were classified as having moderate to high disturbance. Most of the disagreements were only a single category apart, but category assignments among the experts for five of the samples spanned the entire range of condition categories. Expert B did not assign any samples to the undisturbed category, and Expert E assigned only one sample to this category.

The experts used eight types of indicators to assess benthic assemblage condition (Table 4). The

Table 3. Mesohaline category designations of eight experts (A - H). The numbers correspond to four disturbance categories from undisturbed (1) to highly disturbed (4).

Sample	A	B	C	D	E	F	G	H
M01	1	2	2	1	3	3	4	3
M02	3	4	3	3	2	1	1	1
M03	4	4	3	3	3	3	4	3
M04	4	3	2	3	3	3	2	2
M05	4	3	2	3	3	4	3	2
M06	2	2	2	2	3	3	2	2
M07	1	3	1	2	1	1	1	2
M08	4	4	4	4	3	4	4	4
M09	3	2	4	4	4	4	4	4
M10	1	3	1	2	3	2	2	1
M11	3	2	4	3	3	3	3	3
M12	2	2	3	2	3	3	2	2
M13	3	3	3	3	3	3	2	3
M14	2	4	1	2	3	2	3	2
M15	1	3	2	1	3	2	3	2
M16	1	2	1	2	2	1	1	2
M17	3	4	1	3	3	2	2	2
M18	4	4	2	3	3	3	3	2
M19	3	4	2	3	3	3	1	1
M20	3	3	3	3	3	3	3	1

number of taxa was used by all eight experts. Diversity metrics and the proportion of tolerant taxa were each used by a majority of the experts; all other indicators were used by less than half of the experts. The most commonly used tolerant indicator taxa included oligochaetes, capitellid polychaetes, and the polychaete *Streblospio benedicti*. Sensitive indicator taxa included gammarid amphipods, bivalves, and other soft bodied invertebrate phyla (Table 5). Only three of those taxa were used by a majority of the experts; most of the indicator taxa were used by only one or two of the experts.

Tidal Freshwater Samples

The relative sample rankings were even more poorly correlated among experts for the tidal freshwater habitat than for the mesohaline, with an average correlation coefficient of 0.29 and only 52% of the expert pairs significantly correlated (Table 6). The ranks of Expert C were inversely correlated with four of the other experts' ranks. All of the experts' ranks, except C and F, were significantly correlated with the median expert rank (Table 7).

There were no samples for which all of the experts agreed on the condition category, but a majority of the experts agreed on categories for 17 of the 20 (Table 8). There were obvious differences in expert perspectives: Expert B placed all samples into the moderately disturbed category, Expert C did not assign any samples as undisturbed, and Expert D did not assign any samples as highly disturbed; all of the other experts assigned at least one sample to each of the disturbance categories. One of the experts noted that five of the tidal freshwater samples had fauna that were characteristic of submerged aquatic vegetation (SAV; Table 8), with the variability in categorization among experts generally higher for those samples.

The metrics and indicators used by the experts for assessing benthic assemblage condition in the tidal freshwater habitat were similar to those used in the mesohaline habitat (Table 4). Six of the experts used taxa numbers and five of the experts used tolerant taxa proportion to assess benthic condition, while the other indicators were used by less than half of the experts. Tolerant indicator taxa included several species of tubificid oligochaetes (currently considered to be part of the family Naiidae) and chironomids. Sensitive taxa primarily included corophid and gammarid amphipods, and molluscs (Table 5). Each of the experts used a different set of taxa, with none of the indicator taxa being used by more than two experts.

Discussion

The level of agreement among the experts in this study was considerably less than reported in a similar study conducted for higher salinity California assemblages (Weisberg *et al*. 2008). That study showed an average correlation of 0.92 among expert rankings for the polyhaline San Francisco Estuary, whereas the present study showed average correlations of only 0.38 and 0.29 for the mesohaline and tidal freshwater habitats, respectively. Similarly, the pair-wise agreement among experts assigning samples to four disturbance categories averaged 41 and 39% for the mesohaline and tidal freshwater habitats respectively, compared to 70% agreement in the polyhaline study.

Table 4. Indicators used by the experts to evaluate samples. n is the number of experts for each habitat type. The table presents the number of experts that used each indicator as a primary or secondary determinant of condition.

Indicators	Mesohaline (n = 8)		Tidal Freshwater (n = 7)	
	Primary	Secondary	Primary	Secondary
Number of taxa	5	3	2	4
Total abundance	1	3	1	1
Diversity, dominance, evenness metrics	3	2	2	0
Species composition	1	2	0	1
Proportion or dominance of tolerant taxa	3	2	3	2
Proportion or dominance of sensitive taxa	1	2	2	1
Other indicator taxa, higher taxa	0	2	0	3
Life history traits	0	2	0	2

Table 5. Tolerant and sensitive taxa used for evaluation by the experts in each assemblage. n = the number of experts that used each taxon. Note that Tubificidae are currently considered to be part of the family Naiidae.

Mesohaline	n	Tidal Freshwater
Tolerant Taxa		**Tolerant Taxa**
Oligochaeta	5	*Aulodrilus* spp
Capitellidae	4	*Bothrioneurum vejdovskyanum*
Streblospio benedicti	4	*Branchiura sowerbyi*
Grandiderella japonica	3	Chironomidae
Neanthes spp	3	*Chironomus attenuatus*
Eteone lighti	2	*Cryptochironomus* spp
Glycinde armigera	2	*Dero digitata*
Heteromastus sp	2	*Ilyodrilus* spp
Mediomastus spp	2	*Limnodrilus hoffmeisteri*
Nippoleucon hinumensis	2	*Paratanytarsus* sp A
Polydora cornuta	2	*Procladius* sp A
Theora lubrica	2	*Psectrocladius* sp A
Ampelisca abdita	1	*Quistadrilus multisetosus*
Bivalvia	1	*Tanytarsus* sp A
Corbula amurensis	1	*Ablabesmyia* sp A
Macoma spp	1	*Dicrotendipes* spp
Musculista stenhousia	1	Oligochaeta
Nematoda	1	*Polypedilum* sp A
Tubificidae	1	Tubificidae
Sensitive Taxa		**Sensitive Taxa**
Ampelisca abdita	2	Amphipoda
Amphipoda	2	Gammaridea
Gammaridae	2	Corophidae
Mollusca	2	*Hyallela* spp
Planaria, Anthozoa, Nemertea, Ectoprocta, Echiura	2	*Americorophium* spp
Polychaeta	2	Mollusca
Sabaco elongatus	2	*Corbicula fluminea*
Corbula amurensis	1	*Pisidium compressum*
Corophidae	1	*Manayunkia speciosa*
Crustacea	1	*Laonome* spp
Gemma gemma	1	Insecta
Leitoscoloplos sp	1	*Varichaetodrilus angusitpenis*
Leptochelia dubia	1	*Sparganophilus eiseni*
Mya arenaria	1	*Planaria, Nemertea, Odonata, Ephemeroptera, Ostracoda*
Ostracoda	1	
Synidotea laticauda	1	

There are several possible reasons why there was less agreement among experts for the lower salinity habitats. One reason for less agreement is that typically reduced species richness in low salinity habitats (Dauer 1993, Engle *et al*. 1994, Ranasinghe *et al*. in press) may reduce the number of known indicator taxa that could differentiate samples. The polyhaline samples provided to experts in the polyhaline study (Weisberg *et al*. 2008) averaged 31 species, whereas the samples for this study averaged only 14 and 9 species in the mesohaline and tidal freshwater habitats, respectively.

Another possible reason for less agreement may be related to differences in the assemblages' disturbance regimes. The types, scales, frequencies, and magnitudes of habitat disturbance are probably greater in the tidal freshwater (Moyle *et al*. 2010) and mesohaline habitats than in polyhaline assemblages. These disturbances may include osmotic stress, organic enrichment, seasonal freshwater inflows and diversions, channel dredging, shipping traffic, and agricultural discharges. Disturbances probably select for more tolerant taxa and preclude the presence of sensitive ones. Based on available

Table 6. Spearman rank correlation coefficients between expert tidal freshwater sample rankings. Positive, significant ($p < 0.05$) correlations are bold.

	A	B	C	D	E	F
A						
B	**0.571**					
C	-0.332	-0.269				
D	**0.809**	**0.508**	-0.463			
E	**0.460**	**0.492**	0.032	**0.453**		
F	0.070	0.163	0.333	0.018	**0.676**	
G	**0.935**	**0.570**	-0.262	**0.884**	**0.451**	0.056

information in the literature about taxon tolerance or sensitivity 90% of the mesohaline BPJ samples had more (>50%) tolerant than sensitive taxa, and 60% of the tidal freshwater samples had more tolerant than sensitive taxa, compared to 36% tolerant taxa for the polyhaline samples. Elliott and Quintino (2007) termed this the estuarine quality paradox, in which structural community measures are confounded by natural physical stresses, and suggested that greater reliance may need to be placed on functional measures to achieve quality assessments in transitional waters. The apparent reduction of one important type of disturbance, sediment contamination, may have affected the experts' ability to distinguish benthic condition in the low salinity samples because they did not include as wide a sediment contamination gradient as samples for the polyhaline study. The polyhaline BPJ samples had a maximum

Table 7. Spearman correlation coefficients between each expert's tidal freshwater sample rankings and the median rank. Significant ($p < 0.05$) correlations are bold.

A	B	C	D	E	F	G
0.556	**0.574**	-0.284	**0.54**	**0.542**	0.171	**0.595**

Table 8. Tidal freshwater condition category designations of seven experts (A - G). The numbers correspond to four disturbance categories from undisturbed (1) to highly disturbed (4). * indicates that the sample included taxa considered to be indicative of submerged aquatic vegetation (SAV).

Sample	A	B	C	D	E	F	G
T01	4	3	4	3	4	4	4
T02 *	3	3	2	3	2	1	4
T03	3	3	3	3	4	4	4
T04 *	3	3	2	3	2	1	3
T05	2	3	3	2	3	3	3
T06 *	3	3	2	2	2	1	3
T07	2	3	3	2	3	3	1
T08	1	3	4	1	1	1	1
T09	1	3	4	1	1	3	1
T10	2	3	3	2	3	2	2
T11	2	3	3	2	3	2	2
T12	1	3	4	1	1	1	1
T13	2	3	3	2	3	3	2
T14	4	3	4	3	4	3	4
T15 *	3	3	3	2	2	2	3
T16	2	3	3	2	2	2	1
T17 *	3	3	2	3	2	1	3
T18	1	3	4	2	3	3	1
T19	2	3	3	2	2	2	3
T20	3	3	3	2	3	2	3

mERMq of 1.82 (Table 9) compared to the tidal freshwater and mesohaline maxima of less than 0.40. Similarly, there were only two samples that had less than 50% survival in amphipod toxicity tests in this study, compared to several samples with nearly no survival in the polyhaline study (Weisberg *et al*. 2008).

During a debriefing session with the experts following the exercise, it was noted that some experts conducted their assessments based primarily on community metrics, while others used the presence or dominance of specific indicator taxa. This was similar to the polyhaline exercise, and one of the experts asked whether the greater agreement for the polyhaline exercise might be due to better convergence of these two strategies in higher salinity habitats. This possibility was investigated by correlating three community metrics (number of taxa, total abundance, and tolerant taxa) and three indicators used in each habitat (number of amphipod taxa, oligochaete abundance, and *Capitella* abundance). Correlations among these indicators were generally low and not substantially different among any of the habitats, suggesting that the difference in expert agreement among habitats was not attributable to better convergence of the indicator classes in the polyhaline than in the lower salinity assemblages. Another expert suggested that some of the difference may have resulted from the way in which experts scaled their assessments. For instance, one expert categorized every site as disturbed, as he believed that reflected the general condition of the Delta. Another expert with experience in other west coast low salinity habitats believed there was greater taxonomic diversity in other locations and used that experience in his categorizations. Most of the other experts scaled their responses to experiences within the San Francisco Estuary, or to the range of conditions within the BPJ samples. The experts agreed that providing clearer instructions with regard to scaling may have enhanced their agreement, but would not have resolved the underlying differences because scaling only affects the categorical comparison, whereas the rank correlations for this study were also much lower than those in the polyhaline exercise.

The most likely explanation for the poor agreement on benthic condition in mesohaline and tidal freshwater habitats appears to be a lack of common knowledge about the responses of potential indicator organisms to stress and disturbance in low salinity habitats. In the polyhaline assemblage, more than half of the experts agreed on 14 tolerant taxa and an additional 8 sensitive indicator taxa. In contrast, only three indicator taxa were used by more than half of the mesohaline experts and no indicator taxa were agreed to by more than two of the tidal freshwater experts (Table 5). Interestingly, there was even disagreement among some experts as to whether the introduced amphipod *Ampelisca abdita* and the introduced clam *Corbula amurensis*, both abundant taxa in the mesohaline, should be considered tolerant or sensitive indicators. That type of disagreement did not occur in the polyhaline habitat.

Regardless of the reasons, the lower level of agreement among experts makes the development of benthic assessment methods difficult in low salinity habitats. Indices rely on agreement about underlying conceptual models of perturbation effects that indices mathematically capture. The Pearson-Rosenberg (1978) model of benthic response to organic enrichment is often used by index developers

Table 9. Ranges of selected aboitic and benthic variables in the BPJ samples from three San Francisco Estuary assemblages. nm = not measured.

Variable	Tidal Freshwater (n = 20)		Mesohaline (n = 20)		Polyhaline (n = 11)	
	Min	Max	Min	Max	Min	Max
Depth	nm	nm	1.9	6.5	0.1	16
Salinity	nm	0.05	7.9	30.8	22.2	30.8
Percent fines	4	99	20	99.6	31	100
TOC	0.1	12.8	0.51	5.1	0.55	6.04
mERMq	0.005	0.398	0.032	0.357	0.127	1.82
Number of taxa	3	24	3	25	0	55
Total abundance	39	2322	72	5583	0	3489

in high salinity environments as a generalized conceptual model for benthic response to a wide range of disturbances. A similar set of gradient studies that yield an appropriate conceptual construct is lacking for lower salinity environments, which seems to be the necessary next step in developing agreement among experts and providing the foundation for index development in those habitats.

Literature Cited

Alden III, R.W., D.M. Dauer, J.A. Ranasinghe, L.C. Scott and R.J. Llanso. 2002. Statistical verification of the Chesapeake Bay benthic index of biotic integrity. *Environmetrics* 13:473-498.

Blocksom, K.A., J.P. Kurtennbach, D.J. Kemm, F.A. Fulk and S.M. Cormier. 2002. Development and evaluation of the lake macroinvertebrate integrity index (LMI) for New Jersey lakes and reservoirs. *Environmental Monitoring and Assessment* 77:311-333.

Borja, A. and D.M. Dauer. 2008. Assessing the environmental quality status in estuarine and coastal systems: comparing methodologies and indices. *Ecological Indicators* 8:331-337.

Borja, A., I. Muxika and J. Franco. 2003. The application of a Marine Biotic Index to different impact sources affecting softbottom benthic communities along European coasts. *Marine Pollution Bulletin* 46:835-845.

Dauer, D.M. 1993. Biological criteria, environmental health and estuarine macrobenthic community structure. *Marine Pollution Bulletin* 26:249-257.

Dauer, D.M., J.A. Ranasinghe and S.B. Weisberg. 2000. Relationships between benthic community condition, water quality, sediment quality, nutrient loads, and land use patterns in Chesapeake Bay. *Estuaries* 23:80-96.

Dauvin, J.C. 2007. Paradox of estuarine quality: Benthic indicators and indices, consensus or debate for the future. *Marine Pollution Bulletin* 55:271–281.

Dauvin, J.C. and T. Ruellet. 2009. The estuarine quality paradox: is it possible to define an ecological quality status for specific modified and naturally stressed estuarine ecosystems? *Marine Pollution Bulletin* 59:38-47.

Diaz, R.J., M. Solan and R.M. Valente. 2004. A review of approaches for classifying benthic habitats and evaluating habitat quality. *Journal of Environmental Management* 73:165-181.

Elliott, M. and V. Quintino. 2007. The estuarine quality paradox, environmental homeostasis and the difficulty of detecting anthropogenic stress in naturally stressed areas. *Marine Pollution Bulletin* 54:640-645.

Engle, V.D., K.J. Summers and G.G. Gaston. 1994. A benthic index of environmental condition of the Gulf of Mexico Estuaries. *Estuaries* 17:372-384.

Hartig, J.H., M.A. Zarull, T.B. Reynoldson, G. Mikol, V.A. Harris, R.G. Randall and V.W. Cairns. 1997. Quantifying targets for rehabilitating degraded areas of the Great Lakes. *Environmental Management* 21:713-723.

Long, E.R., D.D. MacDonald, S.L. Smith and F.D. Calder. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environmental Management* 19:81-97.

Marques, J.C., F. Salas, J. Patrício, H. Teixeira and J.M. Neto. 2009. Ecological Indicators for Coastal and Estuarine Environmental Assessment - a user guide. WIT Press. Southampton, UK.

Moyle, P.B., W.A. Bennett, W.E. Fleenor and J.R. Lund. 2010. Habitat variability and complexity in the upper San Francisco Estuary. *San Francisco Estuary & Watershed Science* 8(3).

Muxika, I., A. Borja and W. Bonne. 2005. The suitability of the marine biotic index (AMBI) to new impact sources along the European coasts. *Ecological Indicators* 5:19-31.

Pearson, T.H. and R. Rosenberg. 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. *Oceanography and Marine Biology: An Annual Review* 16:229-311.

Pinto, R., J. Patricio, A. Baeta, B.D. Fath, J.M. Neto and J.C. Marques. 2009. Review and evaluation of estuarine biotic indices to assess benthic condition. *Ecological Indicators* 9:1-25.

Ranasinghe, J.A., K.I. Welch, P.N. Slattery, D.E. Montagne, D.D. Huff, H. Lee II, J.L. Hyland, B.

Thompson, S.B. Weisberg, J.M. Oakden, D.B. Cadien and R.G. Velarde. *In Press*. Habitat-related benthic macrofaunal assemblages of bays and estuaries of the western United States. *Integrated Environmental Assessment and Management*.

Schloesser, D.W., T.B. Reynoldson and B.A. Manny. 1995. Oligochaete fauna of western Lake Erie 1961 and 1982 – Signs of sediment quality recovery. *Journal of Great Lakes Research* 21:294-306.

Teixeira, H., A. Borja, S.B. Weisberg, J.A. Ranasinghe, D.B. Cadien, D.M. Dauer, J.C. Dauvin, S. Degraer, R.J. Diaz, A. Grémare, I. Karakassis, R.J. Llansó, L.L. Lovell, J.C. Marques, D.E. Montagne, A. Occhipinti-Ambrogi, R. Rosenberg, R. Sardá, R., L.C. Shaffner and R.G. Velarde. 2010. Assessing coastal benthic macrofauna community condition using best professional judgment - Developing consensus across North America and Europe. *Marine Pollution Bulletin* 60:589-600.

US Environmental Protection Agency (USEPA). 1998. Lake and reservoir bioassessment and biocriteria: technical guidance document. EPA 841-B-98-007. USEPA. Washington, D.C.

USEPA. 2002. Summary of Biological Assessment Programs and Biocriteria Development for States, Tribes, Territories, and Interstate Commissions: Streams and Wadeable Rivers. EPA-822-R-02-048. USEPA, Office of Environmental Information and Office of Water. Washington, DC.

Weisberg, S.B., J.A. Ranasinghe, D.M. Dauer, L.C. Schaffner, R.J. Diaz and J.B. Frithsen. 1997. An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. *Estuaries* 20:149-158.

Weisberg, S.B., B. Thompson, J.A. Ranasinghe, D.E. Montagne, D.B. Cadien, D.M. Dauer, D. Diener, J. Oliver, D.J. Reish, R.G. Velarde and J.Q. Word. 2008. The level of agreement among experts applying best professional judgment to assess the condition of benthic infaunal communities. *Ecological Indicators* 8:389-394.

Acknowledgements

Funding for this study was provided by the California State Water Resources Control Board, the San Francisco Estuary Regional Monitoring Program, and the Southern California Coastal Water Research Project. The authors thank Dan Pickard and John Sandberg of the California Department of Fish and Game's Aquatic Bioassessment Laboratory, who assisted one of the experts with tidal freshwater assemblage evaluation.