Ecological Indicators 12 (2012) 167–173

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Low levels of agreement among experts using best professional judgment to assess benthic condition in the San Francisco Estuary and Delta

Bruce Thompson[a,*], Stephen B. Weisberg[b], Aroon Melwani[a], Sarah Lowe[a], J. Ananda Ranasinghe[b], Donald B. Cadien[c], Daniel M. Dauer[d], Robert J. Diaz[e], Wayne Fields[f], Michael Kellogg[g], David E. Montagne[h], Peter R. Ode[i], Donald J. Reish[j], Peter N. Slattery[k]

[a] San Francisco Estuary Institute, 7770 Pardee Lane, Oakland, CA 94621, USA
[b] Southern California Coastal Water Research Project, 3535 Harbor Blvd., Costa Mesa, CA 92626, USA
[c] Sanitation Districts of Los Angeles County, P.O. Box 4998, Whittier, CA 90607, USA
[d] Department of Biological Sciences, Old Dominion University, Norfolk, VA 23529, USA
[e] Virginia Institute of Marine Science, Gloucester Point, VA 23062, USA
[f] Hydrozoology, Newcastle, CA 95658, USA
[g] City and County of San Francisco, San Francisco, CA 94132, USA
[h] P.O. Box 2004, Penn Valley, CA 95946, USA
[i] California Aquatic Bioassessment Laboratory, Rancho Cordova, CA 95670, USA
[j] Department of Biological Sciences, California State University Long Beach, Long Beach, CA 90840, USA
[k] MossLanding Marine Laboratories, Moss Landing, CA 95039, USA

ARTICLE INFO

Article history:
Received 19 November 2010
Received in revised form 13 March 2011
Accepted 18 March 2011

Keywords:
Best professional judgment
Benthic assemblages
San Francisco Estuary

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 low salinity estuarine and tidal freshwater habitats.

1. 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 prevalent (Marques et al., 2009). Benthic indices summarize 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.

Benthic index development has occurred primarily for polyhaline and euhaline environments. Benthic indices are also widely used in freshwater streams (USEPA, 2002) and in some lakes (Schloesser, 1995; Hartig et al., 1997; USEPA, 1998; Blocksom et al., 2002), but they have yet to be successfully developed for low salinity or tidal freshwater habitats in estuaries. Weisberg et al. (1997) developed separate indices for seven Chesapeake Bay benthic habitats, but found reduced levels of success with decreasing

* Corresponding author.
E-mail address: brucet@sfei.org (B. Thompson).

1470-160X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ecolind.2011.03.020

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, 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 stressor effects on benthic organisms and communities. Two recent papers (Weisberg et al., 2008; Teixeira et al., 2010) address the extent to which experts using best professional judgment (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 lack of success in developing benthic assessments in low salinity habitats presages difficulties in obtaining BPJ consensus in those habitats, largely because they are believed to be more complex habitats where effects on benthos are poorly understood.

The objectives of this study were to determine the levels of expert agreement about benthic condition in the mesohaline and tidal freshwater habitats of the San Francisco Estuary and Sacramento – San Joaquin River Delta. The levels of agreement among experts reported in this study are compared to those achieved in a similar study (Weisberg et al., 2008) previously conducted in the San Francisco Estuary polyhaline (high salinity) habitat, and possible reasons for differences in agreement among habitats are suggested.

2. 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 (moderate salinity) assemblage of the San Francisco Estuary. Seven experts were provided data for 20 samples from the tidal freshwater (equivalent to limnetic in the Venice system of terminology) assemblage of the Sacramento-San Joaquin River Delta. The term assemblage is used to refer to the benthic community that inhabits each salinity class habitat. In the San Francisco Estuary and Delta, benthic assemblages generally coincide with salinity class designations (Thompson et al., 2000).

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 through out the United States (US); three of the experts have direct experience in the San Francisco Estuary and/or Delta. The experts are given letter designations when referenced below; experts A, B, C, D, and H, evaluated samples from both assemblages.

The sites were selected to represent the entire range of geography and sediment conditions within each assemblage. Sediment variables used in the selection process were salinity (PSU), sediment contamination (mean Effects Range-Median quotient (mERMq); Long et al., 1995), percent fine sediments (<63 μm), and percent total organic carbon. Sediment contamination data was not provided to the experts.

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.

Statistical analyses were conducted using SAS (2006) and R (2010). For the rank correlations, tied ranks were given the midpoint value. Correlations between each expert's rankings and the median rank used one expert ranking compared the median rank of all other experts. Factor analysis was conducted using un-transformed variables that contributed to eigenvalues greater than 1.

3. Results

3.1. Mesohaline samples

There was generally poor correlation between the relative rankings of the samples by the experts. The average Spearman's correlation coefficient between experts was 0.38, with only 35.7% of the expert pairs correlated with probabilities below 0.01 (Table 1). The rankings of Expert B were inversely correlated with five of the other expert's rankings. Five of the seven expert's ranks were correlated with the median ranks with probability below 0.01 (Table 2). Thus, the ranks provided by Experts B and H were much different than the others.

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). The average coefficient of variation (CV) among the experts sample categories was 30.2 (range = 9.1–51.8), and the average CV within the each experts' categories was 35.7 (range = 20.6–44.8). The lowest levels of variation were for the most degraded samples (e.g. M08). The highest levels of variation were for samples where experts' categories ranged between one and four. 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 (e.g. M02). 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 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 in their evaluations. Most of the indicator taxa were used by only one or two of the experts.

3.2. Tidal freshwater samples

The relative sample rankings were even more poorly correlated among experts for the tidal freshwater assemblage than for the mesohaline, with an average correlation coefficient of 0.29, and only 28.6% of the expert pairs with probabilities below 0.01 (Table 6). The ranks of Expert I were inversely correlated with four of the other experts' ranks. Five of the seven expert ranks were correlated with median rank with probabilities below 0.01. Experts I and H had much different rankings than the other experts (Table 7).

Table 1
Spearman's correlation coefficients and probabilities (p) between each pair of mesohaline expert's condition categories.

		A	B	C	D	E	F	G
B		0.333						
	p	0.151						
C		0.639	−0.187					
	p	0.002	0.430					
D		0.780	0.493	0.605				
	p	<0.001	0.027	0.006				
E		0.505	−0.040	0.508	0.392			
	p	0.023	0.867	0.024	0.088			
F		0.740	−0.015	0.647	0.651	0.635		
	p	<0.001	0.950	0.003	0.002	0.003		
G		0.451	−0.245	0.478	0.211	0.812	0.680	
	p	0.046	0.298	0.035	0.371	<0.001	0.001	
H		0.032	−0.266	0.309	0.204	0.496	0.369	0.565
	p	0.892	0.258	0.186	0.387	0.026	0.110	0.009

Table 2
Spearman's correlation coefficients and probabilities (p) between each mesohaline expert's sample rankings and the median rank.

		A	B	C	D	E	F	G	H
Median rank		0.734	−0.031	0.577	0.608	0.599	0.738	0.495	0.317
	p	<0.001	0.342	<0.001	<0.001	0.002	<0.001	0.015	0.404

Table 3
Mesohaline category designations of eight experts (A–H), and coefficients of variation (CV) for each expert and sample. The categories correspond to disturbance levels between 1 (undisturbed) and 4 (highly disturbed).

Sample	A	B	C	D	E	F	G	H	CV
M01	1	2	2	1	3	3	4	3	44.7
M02	3	4	3	3	2	1	1	1	51.8
M03	4	4	3	3	3	3	4	3	15.3
M04	4	3	2	3	3	3	2	2	25.7
M05	4	3	2	3	3	4	3	2	25.2
M06	2	2	2	2	3	3	2	2	20.6
M07	1	3	1	2	1	1	1	2	50.4
M08	4	4	4	4	3	4	4	4	9.1
M09	3	2	4	4	4	4	4	4	20.5
M10	1	3	1	2	3	2	2	1	44.5
M11	3	2	4	3	3	3	3	3	17.8
M12	2	2	3	2	3	3	2	2	21.8
M13	3	3	3	3	3	3	2	3	12.3
M14	2	4	1	2	3	2	3	2	38.6
M15	1	3	2	1	3	2	3	2	39.3
M16	1	2	1	2	2	1	1	2	35.6
M17	3	4	1	3	3	2	2	2	37.0
M18	4	4	2	3	3	3	3	2	25.2
M19	3	4	2	3	3	3	1	1	42.8
M20	3	3	3	3	3	3	3	1	25.7
CV	43.9	27.1	44.8	31.6	20.6	35.2	42.1	40.7	

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 samples (Table 8). The average coefficient of variation (CV) among the experts' sample categories was 31.6 (range = 13.9–77.5), and the average CV within the each experts' categories was 32.1 (range = 0–46.8). The lowest levels of variation among experts were for the most degraded samples (e.g. T01). The highest levels of variation were for samples where experts' categories ranged between one and four (e.g. L08). There were obvious differences in expert perspectives: Expert B placed all samples into

Table 4
Indicators used by the experts to evaluate samples. n is the number of experts for each habitat type. The table shows 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 by the experts in their evaluations of the mesohaline and tidal freshwater assemblages. *N* is 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	*N*
Tolerant taxa			
Oligochaeta	5	*Limnodrilus hoffmeisteri*	2
Capitellidae	4	*Dero digitata*	2
Streblospio benedicti	4	*Aulodrilus* spp.	2
Neanthes spp.	3	*Bothrioneurum vejdovskyanum*	2
Grandiderella japonica	3	*Branchiura sowerbyi*	2
Mediomastus spp.	2	*Ilyodrilus* spp.	2
Heteromastus sp.	2	*Quistadrilus multisetosus*	2
Eteone lighti	2	Chironomidae	2
Glycinde armigera	2	*Chironomus attenuatus*	2
Polydora cornuta	2	*Cryptochironomus* spp.	2
Theora lubrica	2	*Paratanytarsus* sp. A	2
Nippoleucon hinumensis	2	*Procladius* sp. A	2
Tubificidae	1	*Psectrocladius* sp. A	2
Bivalvia	1	*Tanytarsus* sp. A	2
Corbula amurensis	1	Oligochaeta	1
Musculista senhousia	1	Tubificidae	1
Macoma spp.	1	*Ablabesmyia* sp. A	1
Ampelisca abdita	1	*Dicrotendipes* spp.	1
Nematoda	1	*Polypedilum* sp. A	1
Sensitive taxa			
Amhipoda	2	Gammaridea	2
Gammaridae	2	*Pisidium compressum*	2
Ampelisca abdita	2	*Manayunkia speciosa*	2
Mollusca	2	Amphipoda	1
Polychaeta	2	Corophidae	1
Sacaco elongatus	2	*Hyallela* spp.	1
Planaria, Anthozoa, Nemertea,	2	*Americorophium* spp.	1
Crustacea	1	Mollusca	1
Corophidae	1	*Corbicula fluminea*	1
Leptochelia dubia	1	*Laonome* spp.	1
Synidotea laticauda	1	Insecta	1
Gemma gemma	1	*Varichaetodrilus angusitpenis*	1
Mya arenaria	1	*Sparganophilus eiseni*	1
Corbula amurensis	1	Planaria, Nemertea, Odonata, Ephemeroptera, Ostracoda	1
Leitoscoloplos sp.	1		
Ostracoda, Ectoprocta, Echiura	1		

Table 6
Spearman's correlation coefficients and probabilities (*p*) between each pair of tidal freshwater expert's condition categories.

		A	B	C	D	H	I
B		0.570					
	p	0.009					
C		0.451	0.492				
	p	0.046	0.029				
D		0.935	0.571	0.460			
	p	<0.001	0.009	0.041			
H		0.056	0.163	0.676	0.070		
	p	0.815	0.493	0.001	0.770		
I		−0.262	−0.269	0.032	−0.332	0.333	
	p	0.264	0.250	0.896	0.153	0.151	
J		0.884	0.508	0.453	0.809	0.018	−0.463
	p	<0.001	0.024	0.047	<0.001	0.940	0.041

the moderately disturbed category, Expert I did not assign any samples as undisturbed, and Expert J 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), but there was similar variability in the categorization of those samples as in all other 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

Table 7
Spearman's correlation coefficients and probabilities (*p*) between each expert's tidal freshwater sample rankings and the median rank.

		A	B	C	D	H	I	J
Median rank		0.595	0.574	0.542	0.556	0.171	−0.284	0.540
	p	<0.001	0.002	<0.001	<0.001	0.047	0.673	<0.001

Table 8
Tidal freshwater category designations of seven experts (A–J), and coefficients of variation for each expert and sample. The categories correspond to disturbance levels between 1 (undisturbed) and 4 (highly disturbed).

Sample	A	B	C	D	H	I	J	CV
T01	4	3	4	4	4	4	3	13.9
T02[a]	4	3	2	3	1	2	3	31.8
T03	4	3	4	3	4	3	3	15.1
T04[a]	3	3	2	3	1	2	3	33.6
T05	3	3	3	2	3	3	2	19.0
T06[a]	3	3	2	3	1	2	2	32.9
T07	1	3	3	2	3	3	2	21.3
T08	1	3	1	1	1	4	1	77.5
T09	1	3	1	1	3	4	1	66.5
T10	2	3	3	2	2	3	2	22.6
T11	2	3	3	2	2	3	2	22.6
T12	1	3	1	1	1	4	1	77.5
T13	2	3	3	2	3	3	2	20.1
T14	4	3	4	4	3	4	3	15.3
T15[a]	3	3	2	3	2	3	2	21.3
T16	1	3	2	2	2	3	2	24.1
T17[a]	3	3	2	3	1	2	3	33.6
T18	1	3	3	1	3	4	2	42.5
T19	3	3	2	2	2	3	2	21.3
T20	3	3	3	3	2	3	2	19.0
CV	46.8	0.0	37.8	39.7	45.7	23.2	31.2	

[a] The presence of SAV noted in the sample.

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 in their evaluations, with none of the indicator taxa being used by more than two experts.

3.3. Relationships between expert classification and sample variables

The ranges of several key abiotic and biological variables measured in the BPJ samples from each assemblage are shown in Table 9. Correlations between the experts' median sample ranking and several key sample variables were evaluated to help understand possible influences on the experts' rankings. In the mesohaline assemblage, salinity and the number of taxa per sample were significantly correlated to the median expert rank (Table 10), suggesting that those two sample variables may have influenced expert opinion on the condition of those samples. Factor analysis showed that those two variables, along with median rank, had the highest Factor 1 loadings, thus covaried among the BPJ samples and accounted for nearly half of the variation among the factors (Table 11). Percent fine sediments and TOC loaded highest on Factor 2 indicating an independent pattern, as did mERMq and total abundance on Factor 3. In the tidal freshwater assemblage, percent fine sediments, mERMq, and total abundance per sample were each significantly correlated with the experts' rankings. Those variables covaried (along with TOC) in the tidal freshwater BPJ samples and accounted for a large majority of the variance, with total taxa forming a separate pattern (Factor 2). These results should be interpreted cautiously, as there were many other abiotic factors that may have influenced the experts that were not quantified.

4. 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.

There are several possible reasons for the lesser agreement among experts for the lower salinity habitats, one of which is that the 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 the common understanding by the experts of the disturbance regimes in each assemblage that affect the benthos. 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 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 mERMq of 1.82 (Table 9) compared to the tidal fresh-

Table 9
Ranges of selected abiotic 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
PSU	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

Table 10
Spearman's correlations of selected abiotic and biological variables with expert median sample ranks.

	PSU (%0)	Fines (%)	TOC (%)	mERMq	N. taxa	Tot. abund
Mesohaline:						
Median sample rank	−0.685*	−0.029	0.280	−0.101	−0.761*	−0.034
Tidal freshwater:						
Median sample rank	–	0.527*	0.338	0.478*	−0.138	−0.603*

* Significant ($p<0.05$, $n=20$).

water 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). Although mERMq and expert ranks were significantly correlated in the tidal freshwater samples (Table 10), the maximum mERMq was considered to be moderately low. However, most of the experts placed at least some of the samples in the highly disturbed category, suggesting that sediment contamination alone may not be a key disturbance factor.

During a debriefing session with the experts following the exercise, some experts explained that they 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 comparing the correlations of 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) among assemblages. Correlations among these indicators were generally low and not substantially different between assemblages, suggesting that the difference in expert agreement among habitats was probably not attributable to better convergence of the indicator classes in the polyhaline than in the lower salinity assemblages.

Table 11
Factor analysis results for the mesohaline and tidal freshwater assemblages.

Variable	Factor 1	Factor 2	Factor 3
Mesohaline:			
N. taxa	0.893	0.020	−0.142
PSU	0.848	−0.100	−0.068
Median expert rank	−0.931	0.174	−0.016
TOC	−0.168	0.817	−0.144
Fines	−0.019	0.787	0.313
mERMq	0.048	0.226	0.820
Tot. abund.	0.177	0.093	−0.690
% of variance	47.9	27	23.7

Variable	Factor 1	Factor 2
Tidal freshwater:		
Median expert rank	0.886	−0.051
mERMq	0.775	0.116
Fines	0.677	0.588
TOC	0.589	0.069
Tot. abund.	−0.774	0.203
N. taxa	−0.112	0.936
% of variance	68.5	31.3

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.

A quantitative determination of why there was lower agreement among the experts could not be made because the components of expert judgment could not be quantified. Each expert used a different set of indicators, and integrated their experience differently, even over the range of condition in the samples. Similarly, only a few of the potentially large number of environmental stressors and disturbances that may exist in the low salinity habitats were quantified, and benthic responses to those are largely unknown. Therefore, no firm conclusions about why there was such low agreement in this study can be made.

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 benthic 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.

Without greater knowledge and better understanding of cause and effect in the low salinity habitats, the development of benthic assessment methods for low salinity habitats will be difficult. Indices rely on agreement about underlying conceptual models of stressor or disturbance effects on the benthos that are captured by index formulation. Responses of benthos to stressors and disturbance in high salinity environments (e.g. Pearson and Rosenberg, 1978), may be better understood and agreed upon than in low salinity habitats, possibly because of the apparent complexity of disturbance regimes in low salinity habitats. More knowledge about the disturbance regimes and effects of multiple disturbances is needed in order to construct appropriate conceptual models upon which experts may agree.

Acknowledgments

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, and Kerry Ritter at SCCWRP for statistical advice.

References

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

Blocksom, K.A., Kurtennbach, J.P., Kemm, D.J., Fulk, F.A., Cormier, S.M., 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., Dauer, D.M., 2008. Assessing the environmental quality status in estuarine and coastal systems: comparing methodologies and indices. Ecological Indicators 8, 331–337.

Borja, A., Muxika, I., Franco, J., 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., Ranasinghe, J.A., Weisberg, S.B., 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., Ruellet, T., 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., Solan, M., Valente, R.M., 2004. A review of approaches for classifying benthic habitats and evaluating habitat quality. Journal of Environmental Management 73, 165–181.

Elliott, M., Quintino, V., 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., Summers, K.J., Gaston, G.G., 1994. A benthic index of environmental condition of the Gulf of Mexico Estuaries. Estuaries 17, 372–384.

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

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

Marques, J.C., Salas, F., Patrício, J., Teixeira, H., Neto, J.M., 2009. Ecological Indicators for Coastal and Estuarine Environmental Assessment—A User Guide. WIT Press, Southampton.

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

Muxika, I., Borja, A., Bonne, W., 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., Rosenberg, R., 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., Patricio, J., Baeta, A., Fath, B.D., Neto, J.M., Marques, J.C., 2009. Review and evaluation of estuarine biotic indices to assess benthic condition. Ecological Indicators 9, 1–25.

R, 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ranasinghe, J.A., Welch, K.I., Slattery, P.N., Montagne, D.E., Huff, D.D., Lee II, H., Hyland, J., Thompson, B., Weisberg, S.B., Oakden, J.M., Cadien, D.B., Velarde, R.G., in press. Habitat-related benthic macrofaunal assemblages of bays and estuaries of the western United States. Integrated Environmental Assessment and Management.

SAS, 2006. SAS Institute, Inc. Cary, NC.

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

Thompson, B., Lowe, S., Kellogg, M., 2000. Results of the Benthic Pilot Study 1994–1997. Macrobenthic assemblages of the San Francisco Bay-Delta and their responses to abiotic factors. Technical Report 39, San Francisco Estuary Institute, Oakland, CA, 40 pp.

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

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

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