The effect of sample area and sieve size on benthic macrofaunal community condition assessments in California enclosed bays and estuaries

Kamille K. Hammerstrom[1], J. Ananda Ranasinghe, Stephen B. Weisberg, John S. Oliver[1], W. Russell Fairey[1], Peter N. Slattery[1] and James M. Oakden[1]

ABSTRACT

Benthic macrofauna are used extensively for environmental assessment, but the area sampled and sieve sizes used to capture animals often differ among studies. Here we sampled 80 sites using three different sized sampling areas (0.1, 0.05, 0.0071 m^2) and sieved those sediments through each of two screen sizes (0.5, 1 mm) to evaluate their effect on number of individuals, number of species, dominance, non-metric multidimensional scaling (MDS) ordination, and benthic community condition indices that are used to assess sediment quality in California. Sample area had little effect on abundance but substantially affected numbers of species, which are not easily scaled to a standard area. Sieve size had a substantial effect on both measures, with the 1-mm screen capturing only 74% of the species and 68% of the individuals collected in the 0.5-mm screen. These differences, though, had little effect on the ability to differentiate samples along gradients in ordination space. Benthic indices generally ranked sample condition in the same order regardless of gear, though the absolute scoring of condition was affected by gear type. The largest differences in condition assessment were observed for the 0.0071-m^2 gear. Benthic indices based on numbers of species were more affected than those based on relative abundance, primarily because we were unable to scale species number to a common area as we did for abundance.

INTRODUCTION

Benthic macrofauna are used extensively as indicators of environmental status, with numerous studies demonstrating predictable responses to various types of natural and anthropogenic stress (Pearson and Rosenberg 1978, Dauer *et al*. 2000, Diaz *et al*. 2004, Muxika *et al*. 2005, Borja *et al*. 2009). Benthos have many characteristics that make them useful indicators, including the potential for high exposure to stress. Because benthic organisms have limited mobility and cannot avoid adverse conditions, they are exposed to accumulated contaminants and low concentrations of oxygen in near-bottom waters. Benthic assessments are often based on benthic indices, which translate community composition into a quality classification evaluating whether samples deviate from expectations for reference conditions (Weisberg *et al*. 1997, Van Dolah *et al*. 1999, Rosenberg *et al*. 2004, Muxika *et al*. 2007, Weisberg *et al*. 2008, Ranasinghe *et al*. 2009). Benthic indices have proven to be accurate and sensitive indicators of the condition of the sediments in which benthos live (Diaz *et al*. 2004, Pinto *et al*. 2009).

The sampling methods used to capture benthic macrofauna in support of sediment condition assessments vary considerably among studies. The most substantial differences are the area sampled by the collection gear and the mesh size of the screen used to sieve the sampled sediment. Gears that sample larger areas tend to collect more species (Ferraro *et al*. 1994, Ferraro *et al*. 2006). Sieves with smaller apertures retain more individuals, particularly juveniles and smaller taxa (Reish 1959, Bachelet 1990, James *et al*. 1995). However, these fuller community descriptions involve higher costs to sort and identify larger numbers of individuals and species.

A number of studies have examined the effects

---

[1] *Moss Landing Marine Laboratories, Moss Landing, CA*

of sampling area and sieve size on benthic macrofauna collections (Ferraro and Cole 2004, Ferraro *et al.* 2006), but these studies have generally focused on benthic macrofaunal composition and descriptive community statistics. What is less clear is how sampling area and sieve size affect ranking of sampling site quality through use of benthic quality indices or multivariate modeling. Of particular interest, the State of California recently adopted sediment quality objectives that are based in large part on assessing condition of the benthic macrofaunal community (Bay and Weisberg in press). However, different sample areas and sieve sizes have traditionally been used to sample benthos in various parts of California and the sensitivity of those assessments to gear type are unknown. Here, we collected samples using six combinations of sampling area and sieve size in four California habitats to determine how differences in species composition and abundance due to these sampling differences affect several indicators of sediment quality.

METHODS

Benthic macrofaunal samples were collected between June 21 and October 26, 2004 from 80 sites in 4 California benthic habitats (Table 1). At each site, three nested samples of differing areas (0.0071, 0.05, and 0.1 m^2) were collected from a single drop of paired 0.05-m^2 Van Veen grabs mounted on a single frame. The sediment from one 0.05-m^2 grab, a 0.0071-m^2 core collected from the other grab, and the remaining sediment from the other grab were sieved separately. The core and intact grab provided 0.0071-m^2 and 0.05-m^2 samples, while data from all three portions combined yielded a 0.1-m^2 sample. Each sediment portion was sieved separately through a 1-mm screen, and the residue that passed through was sieved through a 0.5-mm screen. Abundance data from both screens were combined to represent the >0.5-mm fraction. Materials retained on the screens were preserved in buffered 10% formalin in the field and transferred to 70% ethanol within two

Table 1. Description of sampling sites.

Habitat	N	Water Body	Depth (m)			Salinity (psu)			Fines (%)		
			Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Polyhaline San Francisco Bay	11	San Francisco Bay	7.8	7.7	1.5 - 28.7	31.4	0.5	30.8 - 32.6	-	-	-
Southern California Marine Bays	22	Santa Barbara Harbor (6) Los Angeles Harbor (3) Long Beach Harbor (3) Newport Bay (1) Batiquitos Lagoon (1) Mission Bay (1) San Diego Bay (7)	7.9	6.1	0.5 - 21.3	33.9	0.9	32.7 - 35.5	58.3	33.6	0.7 - 96.8
Mesohaline San Francisco Bay	7	San Francisco Bay	2.9	2.2	1.1 - 7.5	28.4	2.6	24.8 - 32.8	86.6	14.8	56.6 - 94.4
Estuaries and Wetlands	40	Arcata Bay (4) Humboldt Bay (4) Tomales Bay (4) Drakes Estero (2) Santa Cruz Harbor (6) Elkhorn Slough (1) Moss Landing Harbor (6) Monterey Harbor (6) Morro Bay (1) Port San Luis (6)	3.8	2.8	0.5 - 10.0	33.9	2.4	25.4 - 36.7	38.5	35.6	1.7 - 99.2

weeks. Animals were sorted from the detritus, identified to the lowest possible taxon, and counted.

Sample area and sieve size effects were evaluated by habitat for three types of assessment measures. First were traditional community measures, including the total numbers of species, total numbers of individuals and Simpson's dominance index (Simpson 1949), as well as numbers of crustacean, mollusc, and polychaete individuals. Two way analysis of variance (ANOVA) for main effects of sieve size and sample area was used to examine within-habitat differences, with the dependent variables, number of individuals, scaled to a standard area (0.1 m^2). Data were log-transformed prior to analysis to meet assumptions of normality and homogeneity of variance, where necessary.

The second approach evaluated whether the sampling sites maintained similar relative positions in multivariate space when using data from the different sample areas and sieve sizes. Permutational analysis of variance (PERMANOVA) was used to evaluate differences due to gear type in non-metric multidimensional scaling (MDS) ordinations of species abundance data, using PRIMER Version 6 software (Anderson *et al*. 2008). Permutational methods were also used to conduct pairwise comparisons among sample areas and between sieve sizes (Anderson *et al*. 2008). Bray-Curtis dissimilarity matrices of fourth-root transformed abundances standardized to 1 m^2 (Anderson *et al*. 2008) were used for these analyses. Separate analyses were conducted for each habitat.

The third approach evaluated sample area and sieve size effects on four indices of benthic community condition that are part of California's sediment quality objectives: the Benthic Response Index (BRI; Smith *et al*. 2001, 2003), the River Invertebrate Prediction and Classification System (RIVPACS; Wright *et al*. 1993, Van Sickle *et al*. 2006), the Relative Benthic Index (RBI; Hunt *et al*. 2001), and the Index of Biotic Integrity (IBI; Thompson and Lowe 2004). The BRI is a measure of the abundance-weighted pollution tolerance of species present in a sample, while RIVPACS is the ratio of observed species to those expected in undisturbed reference samples in similar habitats. The RBI and IBI are multi-metric indices based on community parameters, such as abundance, number of species, and number of individuals in selected indicator taxonomic groups. Ranasinghe *et al*. (2009) describe calculation of these indices. Application of the indices was limited to habitats for which they were previously calibrated; all four indices were previously calibrated for Southern California marine bays and polyhaline central San Francisco Bay (Ranasinghe *et al*. 2009), the IBI was also calibrated for mesohaline San Francisco Bay (Thompson and Lowe 2004) and the RBI was calibrated for all four habitats (Barnett *et al*. 2007).

The effects of gear on the indices were evaluated in two ways. The first was to assess changes in magnitude of the index value. The second was to determine how any changes in index values affected the assessment of condition within the context of California's sediment quality objectives which classifies sediments into four condition assessment categories: 1) Reference - a community that would occur at a reference site for that habitat; 2) Low disturbance - a community that exhibits some indication of stress, but within measurement variability of reference condition; 3) Moderate disturbance - a community that exhibits evidence of physical, chemical, natural or anthropogenic stress; and 4) High disturbance - a community exhibiting a high magnitude of stress. Samples in categories 3 and 4 are considered to be in poor condition, while categories 1 and 2 are considered to be in good condition.

RESULTS

There were no significant differences in the total numbers of individuals collected as a function of sampled area (after standardization to number per m^2; Table 2). There were also no significant differences in numbers of crustacean, mollusc, and polychaete individuals among the three sample areas for any habitats (Table 2). There was, however, a general pattern of greater number of individuals in the smallest sample area.

In contrast, sieve size had a large effect on total numbers of individuals, with 1-mm sieves containing an average of only 68% (53 - 83%) of the individuals collected on 0.5-mm sieves (Figure 1). Screen size also significantly affected total numbers of individuals in all three major taxa in all habitats except in polyhaline Central San Francisco Bay. Crustacean abundance was most affected by sieve size, with significant differences in three of the four habitats (Table 2).

The number of species sampled was significantly affected by both sample area and screen size. The 1-mm screen captured an average of only 74% (58 -

Table 2. P-values from two-way ANOVAs for effect of sample area (top) and sieve size (bottom). Interaction terms are not shown, but none were significant. The number (N) of individuals (Ind), crustaceans (Crust), molluscs (Moll), and polychaetes (Poly) were scaled to 0.1 m^2, while the number of species (Spp) was not. RIVPACS, BRI, and IBI values were only calculated for habitats for which indices had been previously calibrated.

Main Effects	Habitat	N	Spp	Ind	Dom	Crust	Moll	Poly	RBI	RIV PACS	BRI	IBI
Sample Area	Estuaries & Wetlands	40	< 0.01	0.06	0.87	0.89	0.98	0.82	< 0.01			
	SoCal Marine Bays	22	< 0.01	0.37	0.31	0.63	0.99	0.55	< 0.01	< 0.01	0.11	0.04
	Polyhaline Central SFB	11	< 0.01	0.92	0.87	0.64	0.93	0.67	< 0.01	< 0.01	0.09	< 0.01
	Mesohaline SFB	7	< 0.01	0.13	0.37	0.83	0.92	0.93	< 0.01			< 0.01
Sieve Size	Estuaries & Wetlands	40	< 0.01	0.01	0.09	0.02	0.03	0.01	< 0.01			
	SoCal Marine Bays	22	< 0.01	0.05	0.33	0.05	0.13	0.11	< 0.01	0.21	0.93	0.75
	Polyhaline Central SFB	11	0.04	0.53	0.30	0.52	0.46	0.54	0.06	0.15	0.11	0.82
	Mesohaline SFB	7	< 0.01	0.02	0.01	0.02	0.85	0.04	< 0.01			0.12

81%) of the species captured by the 0.5-mm screen (Figure 2). While sample area and sieve size both strongly affected the numbers of species captured, neither affected the identity of dominant species. The top ten species in each habitat were generally the same for all gear combinations (Table 3), although there were a few exceptions. The most prominent exception was the high abundance of the amphipod *Ampelisca abdita* in San Francisco Bay samples sieved through 0.5-mm screens, compared to its near absence when the same samples were sieved through 1.0-mm screens.

Simpson's dominance was generally higher for 1-mm than 0.5-mm sieves (Figure 3). However, dominance was significantly affected by sieve size only in the mesohaline San Francisco Bay habitat (Table 2), where large numbers of the numerically dominant amphipod Ampelisca abdita passed through 1-mm screens and were captured only in 0.5-mm sieves. There were no significant differences in dominance among the sample areas in any habitat (Table 2).

PERMANOVA indicated significant effects of gear area and screen size on community clustering (Table 4), but the effects were small as samples from the four habitats segregated similarly in all MDS plots (Figure 4). Sample area comparisons revealed the largest differences between 0.0071-m^2 and 0.1-m^2 samples in southern California marine bays and estuaries and wetlands habitats. Sieve size significantly affected community clusters for the estuaries and wetlands and mesohaline San Francisco Bay habitat.

Of the four benthic indices evaluated, the BRI was the only one which was unaffected by either gear or sieve size (Figure 5; Table 2). The RBI was most affected, decreasing significantly with decreasing sample area and with the biggest change occur-

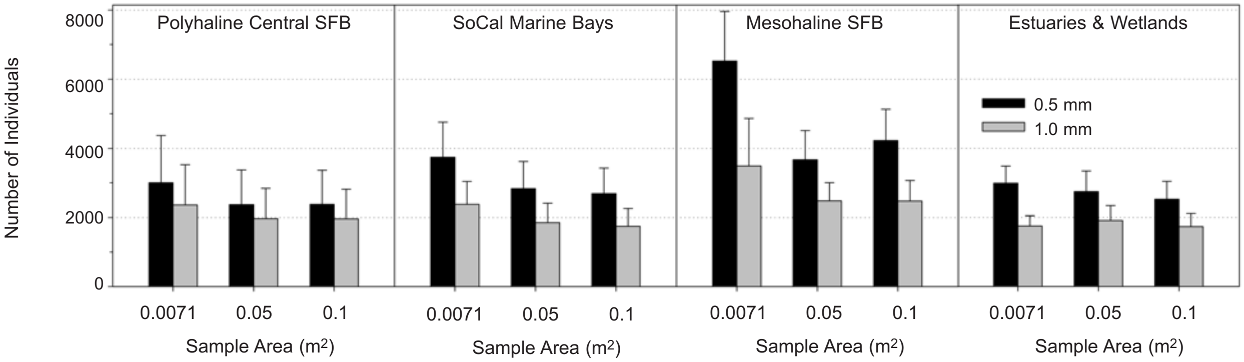

Figure 1. Mean number of individuals collected (per 0.1 m^2) as a function of screen size, sample area, and habitat class (bars represent standard error). N shown in Table 1.

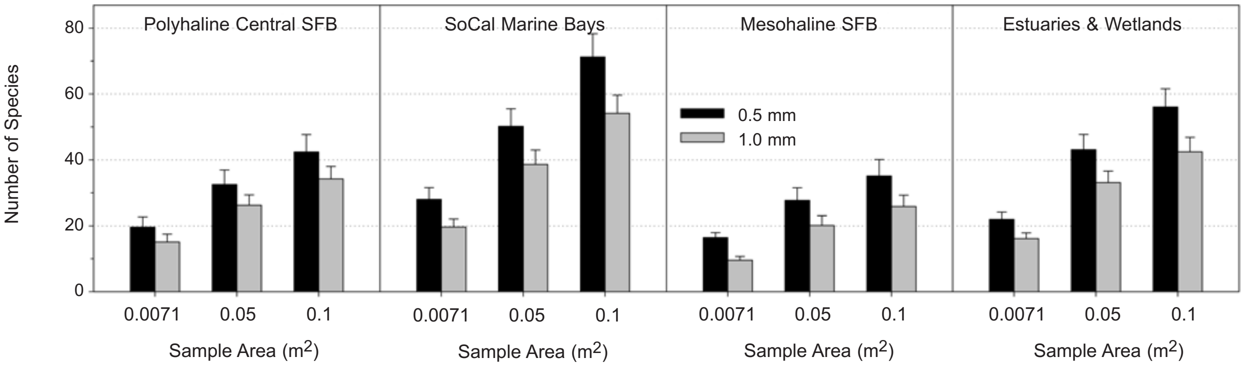

Figure 2. Number of species per sample as a function of screen size, sample area, and habitat class.

ring between the 0.0071-m^2 and 0.05-m^2 gears. The index was also significantly lower for the 1 mm screen size in all four habitats (Figure 5). RIVPACS also decreased with decreasing gear area, with the biggest change between 0.0071-m^2 and 0.05-m^2, but unlike the RBI, sieve size did not affect RIVPACS.

Similarly, the IBI was significantly affected by sample area but not by sieve size. Each index showed the same pattern of change among sample areas and sieve sizes, with the exception of the IBI, which varied (Figure 5).

The frequency with which indices shifted among categories with changing gear was comparable to their relative performance in the ANOVA. For the BRI, we found that less than 12.5% of the samples were classified differently from the classification when using the gear for which the index was calibrated in that habitat, which is comparable to the difference in categorization when comparing two replicates using the same gear (Table 5). In contrast, up to 70% of the samples were classified differently as a function of gear for the RBI. For all of the indices, the 0.0071-m^2 sample area gear had the greatest effect on index categorization (Table 5). Most categorization changes were small and samples that changed were classified into adjacent categories rather than large classification changes into non-adjacent categories (Tables 5 and 6). The highest percentages of large changes were observed for 0.0071-m^2 samples for the RBI, RIVPACS and RBI (Table 6). Except for RBI status categorizations for 0.0071-m^2 samples, most samples did not change status between "good" and "poor" condition as a function of sampling gear (Table 7).

There was pattern to the direction of category change (Table 8). For the IBI, most changes from calibration gear yielded classifications indicating increased disturbance. For the BRI, RIVPACS and RBI, changing to 0.0071-m^2 gear also resulted in classifications indicating increased disturbance, while changing to larger gear and the smaller sieve size consistently decreased disturbance assessments (Table 8).

DISCUSSION

Our findings are consistent with several previous studies that found abundance (after normalization to a common area) is affected more by sieve size than by area of the sample gear (Reish 1959, Ferraro *et al*. 1989, Ferraro *et al*. 1994, James *et al*. 1995, Schlacher and Wooldridge 1996, Ferraro and Cole 2004, Ferraro *et al*. 2006). This results mostly because the smaller sieve size allows capture of juveniles and species that aren't large enough to be retained on the 1-mm screen. In contrast, species richness was more affected by sampling area of the gear than by sieve size. This difference is because the number of individuals is linearly related to sample area and can be multiplicatively adjusted, whereas the relationship between number of species and sample area is nonlinear (Preston 1948, Connor and McCoy 1979, Rey *et al*. 1982) and varies among habitats; thus species richness cannot be normalized for as easily as abundance.

Similar to previous studies, gear and sieve size had less effect on description of community composition than it did on abundance and species number. In several earlier studies, the dominant species were relatively unaffected by sample area and sieve size (Ferraro and Cole 2004) and the clustering of samples in ordination space was affected only to a small degree by sample area or sieve size (James *et al*. 1995, Ferraro *et al*. 2006). Since the development and calibration of indices of condition varies signifi-

Table 3. Rank order of species abundance in samples from each habitat class as a function of sample area and sieve size. The top ten species from the 0.1-m^2 gear and 0.5-mm sieve are listed in the first column for comparison with other gear combinations.

Habitat	Species	Gear Area					
		0.1 m^2		**0.05 m^2**		**0.0071 m^2**	
		0.5 mm	1.0 mm	0.5 mm	1.0 mm	0.5 mm	1.0 mm
Estuaries and Wetlands	*Exogone lourei*	1	2	2	2	4	5
	Oligochaeta	2	1	1	1	2	2
	Paracorophium sp	3	5	8	8	1	1
	Leptochelia dubia	4	4	3	3	5	4
	Nutricola confusa	5	3	4	4	6	3
	Capitella capitata complex	6	7	5	5	9	7
	Monocorophium insidiosum	7	6	6	6	7	9
	Armandia brevis	8	8	7	7	3	6
	Nutricola ovalis	9				8	
	Mediomastus sp	10	9	9	9	10	8
SoCal Marine Bays	*Mediomastus* sp	1	1	1	1	1	1
	Pseudopolydora paucibranchiata	2	3	2	3	2	3
	Nutricola confusa	3	2	4	2	3	2
	Exogone lourei	4	4	3	4	5	6
	Leptochelia dubia	5	5	6	6	4	4
	Oligochaeta	6	6	5	5	7	5
	Nutricola ovalis	7		7		6	
	Cossura sp A	8	7	8	7	10	7
	Grandidierella japonica	9	8	9	8	9	8
	Monocorophium acherusicum	10		10			
Polyhaline Central SFB	*Monocorophium acherusicum*	1	1	1	1	1	1
	Ampelisca milleri	2	2	2	2	5	4
	Molgula sp	3	3	3	3	2	2
	Ampelisca abdita	4		4		3	
	Diadumene sp	5	4	5	4	6	5
	Pseudopolydora paucibranchiata	6	5	9	10	7	6
	Oligochaeta	7	6				
	Euchone limnicola	8		6	9		
	Grandidierella japonica	9	10	8	7	8	7
	Cirriformia sp	10	7	10	6	4	3
Mesohaline SFB	*Ampelisca abdita*	1		2		1	
	Ampelisca milleri	2	1	1	1	3	2
	Nutricola confusa	3	2	3	2	2	1
	Nippoleucon hinumensis	4	4	4	4	6	4
	Potamocorbula amurensis	5	3	5	3	4	3
	Streblospio benedicti	6	5	6	6	5	5
	Ampelisca sp.	7	6	7	5		8
	Euchone limnicola	8	7	8	8	9	6
	Heteromastus filiformis	9	9	10	10	8	10
	Grandidierella japonica	10	10		9	7	9

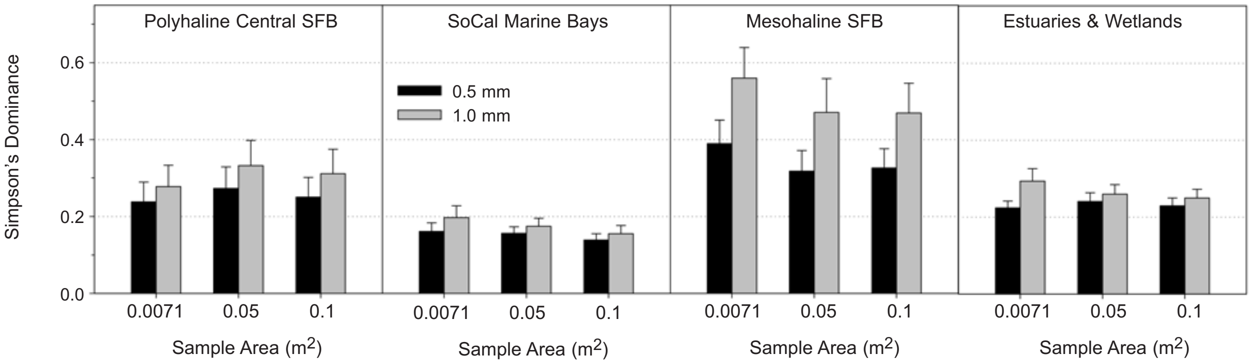

Figure 3. Simpson's dominance as a function of screen size, sample area, and habitat class.

cantly among different types of communities (Ranasinghe *et al*. 2009), an early step in index use is to perform cluster analyses to separate the community types (Ranasinghe *et al*. in press), so that indices can be calibrated from each community (Ranasinghe *et al*. 2009). Our study also showed that sample area and sieve size had little effect upon measurement of community segregation as illustrated in MDS plots (Figure 4).

The effect of gear area and sieve size on benthic index values, though, depended on the type of index. The BRI was least affected (Figure 5), probably because it is based primarily on species composition, which was not affected much by gear and sieve size. In contrast, the RBI was most affected because it is based largely on overall numbers of species and numbers of species within selected indicator taxonomic groupings, which were most affected by gear and sieve size. Even the RBI values, though, were significantly correlated across gear types (Figure 5), which suggests that application of these indices only requires a recalibration to identify gear specific threshold values for appropriate condition assessment. Of the community, multivariate and benthic index measures we investigated, benthic indices such as the RBI and the IBI, which are based on numbers of species, showed the clearest need for gear specific calibration. However, if we were able to scale species number to a standard area (Table 2), we would expect these indices to have much greater agreement among the three gear types.

Ultimately, the choice of gear depends on the desired goal of sampling (Bachelet 1990, Schlacher and Wooldridge 1996). The larger gear area and smaller sieve size provide a more thorough picture of the community, which can be highly desirable for a complete community description. In many benthic surveys the primary goal is not to assess the relative condition or health of the community, but to describe the structure for comparison to other places and times. A 1-mm sieve often misses too many species

Table 4. PERMANOVA results with pairwise comparison results indicated by letters; sample areas with the same letters are not significantly different.

Habitat	N	Sieve Size Pseudo-F	Sieve Size p	Gear Area Pseudo-F	Gear Area p	Gear Area (m^2) Comparison		
						0.0071	0.05	0.1
Polyhaline Central SFB	66	0.89	0.527	0.99	0.454			
SoCal Marine Bays	132	0.97	0.500	1.44	0.054	A	AB	B
Mesohaline SFB	42	4.19	0.004	1.17	0.304			
Estuaries & Wetlands	240	2.20	0.003	1.91	0.003	A	B	B

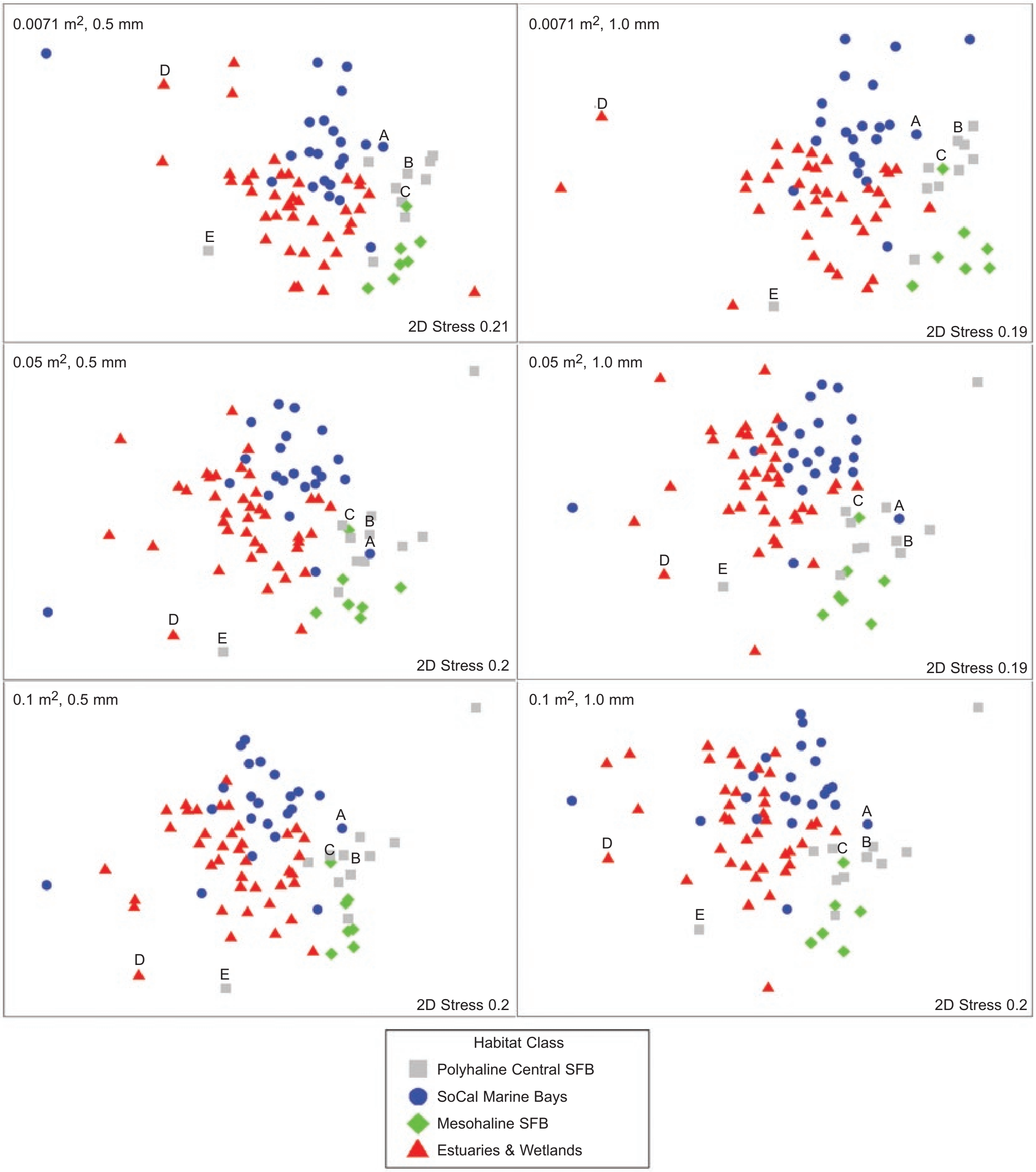

Figure 4. Multidimensional scaling plots for each combination of gear and sieve size. Letters A through E indicate the position of five stations representing each of the four habitat types and illustrate station shifts among the MDS plots.

and individuals to provide an accurate description. However, the amount of processing time is largely proportional to the amount of area sampled and about doubles for processing a 0.5-mm sieve compared to that for a 1.0-mm sieve. Given that we observed little effect on community ordination or indices such as the BRI, the additional detail gained by the larger gear and smaller sieve may be desirable but unnecessary in assessing a sample's condition relative to a reference or healthy community.

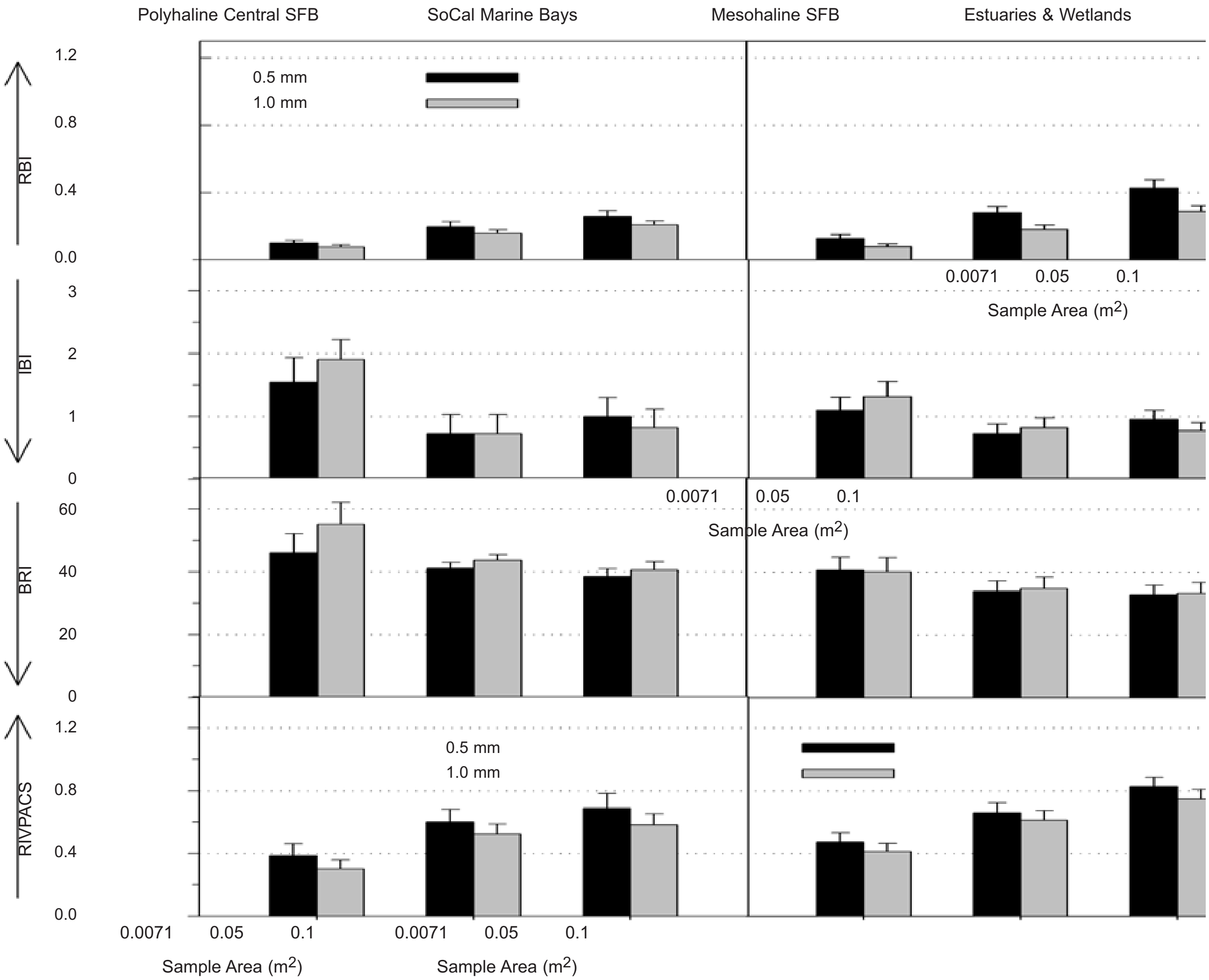

Figure 5. Effect of screen size and sample area on four benthic indices. The arrows indicate the direction of positive change in the community for each index.

Table 5. Percentage of samples for which condition categories changed as a function of gear area and sieve size differerences from those originally used to calibrate the index. Condition categories: (1) Reference, (2) Low disturbance, (3) Moderate disturbance, and (4) High disturbance. Indices were calibrated on data from 0.1-m^2 grabs sieved through 1-mm screens, except in San Francisco Bay, where calibration data were from 0.05-m^2 grabs and sieved through 0.5-mm screens.

Gear Area (m^2)	Sieve Size (mm)	BRI	RIVPACS	RBI	IBI
0.1	0.5	6.3	10.0	31.3	13.8
	1.0	11.1	11.1	33.3	22.2
0.05	0.5	8.1	11.3	22.6	8.1
	1.0	2.5	13.8	36.3	7.5
0.0071	0.5	8.8	22.5	58.8	21.3
	1.0	12.5	27.5	70.0	26.3

Table 6. Percentage of samples with large changes into non-adjacent condition categories as a function of gear area and sieve size differences from those originally used to calibrate the index. Condition categories: (1) Reference, (2) Low disturbance, (3) Moderate disturbance, and (4) High disturbance. Indices were calibrated on data from 0.1-m^2 grabs sieved through 1-mm screens, except in San Francisco Bay, where calibration data were from 0.05-m^2 grabs and sieved through 0.5-mm screens.

Gear Area (m^2)	Sieve Size (mm)	BRI	RIVPACS	RBI	IBI
0.1	0.5	0.0	1.3	5.0	2.5
	1.0	5.6	0.0	0.0	11.1
0.05	0.5	0.0	4.8	8.1	0.0
	1.0	0.0	2.5	6.3	0.0
0.0071	0.5	2.5	8.8	20.0	7.5
	1.0	1.3	10.0	40.0	10.0

Table 7. Percentage of samples for which status classification did not change (from "disturbed" to "undisturbed", or vice versa) as a function of gear area and sieve size differences from those originally used to calibrate the index. Indices were calibrated on data from 0.1-m^2 grabs sieved through 1-mm screens, except in San Francisco Bay, where calibration data were from 0.05-m^2 grabs and sieved through 0.5-mm screens.

Gear Area (m^2)	Sieve Size (mm)	BRI	RIVPACS	RBI	IBI
0.1	0.5	96.3	93.8	86.3	93.8
	1.0	88.9	100.0	94.4	83.3
0.05	0.5	100.0	90.3	85.5	98.4
	1.0	100.0	93.8	83.8	98.8
0.0071	0.5	95.0	87.5	72.5	86.3
	1.0	92.5	82.5	55	83.8

LITERATURE CITED

Anderson, M.J., R.N. Gorley and K.R. Clarke. 2008. PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods. PRIMER-E. Plymouth, UK.

Bachelet, G. 1990. The choice of a sieving mesh size in the quantitative assessment of marine macrobenthos: a necessary compromise between aims and constraints. *Marine Environmental Research* 30:21-35.

Barnett, A.M., S.M. Bay, K.J. Ritter, S.L. Moore and S.B. Weisberg. 2007. Sediment quality in California bays and estuaries. Technical Report 522. Southern California Coastal Water Research Project. Costa Mesa, CA.

Bay, S.M. and S. Weisberg. In press. A framework for interpreting sediment quality triad data. *Integrated Environmental Assessment and Management*.

Borja A, J.A. Ranasinghe and S.B. Weisberg. 2009. Assessing ecological integrity in marine waters using multiple indices and ecosystem components: Challenges for the future. *Marine Pollution Bulletin* 59:1-4.

Clarke, K.R. and R.M. Warwick. 2001. Change in Marine Communities: An Approach to Statistical Analysis and Interpretation. PRIMER-E. Plymouth, UK.

Connor, E.F. and E.D. McCoy. 1979. The statistics and biology of the species-area relationship. *American Naturalist* 113:791-833.

Table 8. Percentages of samples assessed in categories with increasing (I) or decreasing (D) disturbance, for samples where condition categories changed as a function of gear area and sieve size differences from those originally used to calibrate the index.

Gear Area (m^2)	Sieve Size (mm)	Direction	BRI	RIVPACS	RBI	IBI
0.1	0.5	D	100	100	100	9
		I	0	0	0	91
	1.0	D	100	50	33	0
		I	0	50	67	100
0.05	0.5	D	60	29	57	20
		I	40	71	43	80
	1.0	D	0	0	0	0
		I	100	100	100	100
0.0071	0.5	D	14	0	0	6
		I	86	100	100	94
	1.0	D	10	0	0	0
		I	90	100	100	100

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.

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.

Ferraro, S.P. and F.A. Cole. 2004. Optimal benthic macrofaunal sampling protocol for detecting differences among four habitats in Willapa Bay, Washington, USA. *Estuaries* 27:1014-1025.

Ferraro, S.P., F.A. Cole, W.A. DeBen, and R.C. Swartz. 1989. Power-cost efficiency of eight macrobenthic sampling schemes in Puget Sound, Washington, USA. *Canadian Journal of Fisheries and Aquatic Sciences* 46:2157-2165.

Ferraro, S.P., F.A. Cole, and A.R. Olsen. 2006. A more cost-effective EMAP benthic macrofaunal sampling protocol. *Environmental Monitoring and Assessment* 116:275-290.

Ferraro, S.P., R.C. Swartz, F.A. Cole and W.A. De Ben. 1994. Optimum macrobenthic sampling protocol for detecting pollution impacts in the Southern California Bight. *Environmental Monitoring and Assessment* 29:127-153.

Hunt, J.W., B.S. Anderson, B.M. Phillips, R.S. Tjeerdema, K.M. Taberski, C.J. Wilson, H.M. Puckett, M. Stephenson, R. Fairey and J.M. Oakden. 2001. A large-scale categorization of sites in San Francisco Bay, USA, based on the sediment quality triad, toxicity identification evaluations, and gradient studies. *Environmental Toxicology and Chemistry* 20:1252-1265.

James, R.J., M.P.L. Smith and P.G. Fairweather. 1995. Sieve mesh-size and taxonomic resolution needed to describe natural spatial variation of marine macrofauna. *Marine Ecology - Progress Series* 118:187-198.

Muxika, I., A. Borja and J. Bald. 2007. Using historical data, expert judgment and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive. *Marine Pollution Bulletin* 55:16-29.

Muxika, I., A. Borja and W. Bonne. 2005. The suitability of the marine biotic index (AMBI) to new impact sources along 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.

Preston, F.W. 1948. The commonness, and rarity, of species. *Ecology* 29:254-283.

Ranasinghe, J.A., S.B. Weisberg, R.W. Smith, D.E. Montagne, B. Thompson, J.M. Oakden, D.D. Huff, D.B. Cadien, R.G. Velarde and K.J. Ritter. 2009. Calibration and evaluation of five indicators of benthic community condition in two California bay and estuary habitats. *Marine Pollution Bulletin* 59:5-13.

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.*

Reish, D.J. 1959. A discussion of the importance of the screen size in washing quantitative marine bottom samples. *Ecology* 40:307-309.

Rey, J.R., D.R. Strong and E.D. McCoy. 1982. On overinterpretation of the species-area relationship. *American Naturalist* 119:741-743.

Rosenberg, R., M. Blomqvist, H.C. Nilsson, H. Cederwall and A. Dimming. 2004. Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive. *Marine Pollution Bulletin* 49:728-739.

Schlacher, T.A. and T.H. Wooldridge. 1996. How sieve mesh size affects sample estimates of estuarine benthic macrofauna. *Journal of Experimental Marine Biology and Ecology* 201:159-171.

Simpson, E.H. 1949. Measurement of diversity. *Nature* 163:688.

Smith, R.W., M. Bergen, S.B. Weisberg, D.B. Cadien, A. Dalkey, D.E. Montagne, J.K. Stull and R.G. Velarde. 2001. Benthic response index for assessing infaunal communities on the southern California mainland shelf. *Ecological Applications* 11:1073-1087.

Smith, R.W., J.A. Ranasinghe, S.B. Weisberg, D.E. Montagne, D.B. Cadien, T.K. Mikel, R.G. Velarde and A. Dalkey. 2003. Extending the southern California Benthic Response Index to assess benthic condition in bays. Technical Report 410. Southern California Coastal Water Research Project. Westminster, CA.

Thompson, B. and S. Lowe. 2004. Assessment of macrobenthos response to sediment contamination in the San Francisco Estuary, California, USA. *Environmental Toxicology and Chemistry* 23:2178-2187.

Van Dolah, R.F., J.L. Hyland, A.F. Holland, J.S. Rosen and T.R. Snoots TR. 1999. A benthic index of biological integrity for assessing habitat quality in estuaries of the southeastern USA. *Marine Environmental Research* 48:269-283.

Van Sickle, J., D.D. Huff and C.P. Hawkins. 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. *Freshwater Biology* 51:359-372.

Weisberg, S.B., J.A. Ranasinghe, L.C. Schaffner, R.J. Diaz, D.M. Dauer 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.R. Diener, J.S. 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.

Wright J.F., M.T. Furse and P.D. Armitage. 1993. RIVPACS: A technique for evaluating the biological water quality of rivers in the UK. *European Water Pollution Control* 3:15-25.

ACKNOWLEDGEMENTS

The authors would like to thank Walter Nelson and the United States Environmental Protection Agency's Environmental Monitoring and Assessment Program for data, as well as Chris Beegan and the California State Water Resources Control Board for funding to conduct the analyses.