Determining Benthic Invertebrate Community Condition in Embayments

Southern California Coastal Water Research Project
January 21, 2008

INTRODUCTION

Research in California embayments has shown that the use of a combination of benthic indices provides a more accurate description of benthic invertebrate community condition than does the use of a single index (Ranasinghe *et al*. 2007). This document describes the steps necessary to calculate four benthic indices:

1. Index of Biotic Integrity (IBI);
2. Relative Benthic Index (RBI);
3. Benthic Response Index (BRI); and
4. River Invertebrate Prediction and Classification System (RIVPACS).

Details about the history, background, and development of the indices and literature citations are provided in Ranasinghe *et al*. (2007). Each index assesses benthic index condition of a sample as one of four condition categories:

- Reference: A community that would occur at a reference site for that habitat;
- Slight Disturbance: A community that may exhibit some indication of stress, but is within measurement variability of reference condition;
- Moderate Disturbance: A community that exhibits clear evidence of physical, chemical, natural, or anthropogenic stress;
- High Disturbance: A community exhibiting a high magnitude of stress.

The steps necessary to determine the benthic community condition for a sample are:

1. Gathering data;
2. Calculating four benthic indices, and comparing benthic index values to threshold values to determine condition categories; and
3. Integrating the individual index results to determine the benthic community condition classification.

Details of the steps to determine the benthic community condition category for a sample follow.

STEP 1: PREPARE DATA

The raw data needed for the analyses include the abundance of each species (or lowest possible identification level or taxon) and station depth, latitude, and longitude. Each taxon should be identified to the appropriate level in keeping with the benthic macrofauna species list for the relevant habitat. When new taxa are encountered, the nomenclature and level of taxonomy should follow the species list, as far as possible.

STEP 2: CALCULATE BENTHIC INDICES AND DETERMINE CONDITION CATEGORIES

Each benthic index must be calibrated to the natural assemblage characteristic of the sample site habitat. Several different assemblages are present in California embayments. The nature of the expected assemblage is determined by habitat factors, such as salinity, grain size, bottom depth, latitude, and longitude. This document describes the calculation of benthic indices for the assemblages that are characteristic of two habitat types: Southern California Marine Bays and Polyhaline Central San Francisco Bay. Although the same four benthic indices are calculated for samples from each habitat, the metrics included in the RBI and IBI differ slightly, as do the species tolerance values for the BRI and habitat variables and lists of taxa for RIVPACS. Details of the metrics included in each habitat are presented in Table 1 and Table 8. Species lists that include habitat specific species tolerance values for the BRI and RIVPACS reference taxa are provided separately for each habitat.

A. Southern California Marine Bays

Table 1 presents details of the metrics included for calculation of index values in Southern California Marine Bays. Instructions for calculating each of the indices and descriptions of the species list variables follow.

Table 1. Benthic indicator metrics in Southern California Marine Bays. Metrics with superscript letters are included in both the RBI and IBI.

Index	Metric	Use
IBI	Total number of taxa[a]	All taxa
	Number of mollusc taxa[b]	Molluscs
	Notomastus sp. abundance	*Notomastus* sp
	Abundance percentage of sensitive taxa	IBISensitive = S[a]
RBI	Total number of taxa[a]	All taxa
	Number of mollusc taxa[b]	Molluscs
	Number of crustacean taxa	Crustaceans
	Number of crustacean individuals	Crustaceans
	Abundance of *Monocorophium insidiosum*	*Monocorophium insidiosum*
	Abundance of *Asthenothaerus diegensis*	*Asthenothaerus diegensis*
	Abundance of *Goniada littorea*	*Goniada littorea*
	Presence of *Capitella capitata* complex	*Capitella capitata* Cmplx
	Presence of Oligochaeta	Oligochaeta
BRI	Abundance-weighted average tolerance score	ToleranceScore
RIVPACS	Observed to expected (O/E) ratio for number of RIVPACS reference taxa.	Instructions for calculating O/E Ratio using SAS Software (Appendix A) or the Utah State University web site (Appendix B).

[a] Defined in species list (Table 7).

(i) Index of Biotic Integrity (IBI) and IBI condition category

The IBI compares the values of four different metrics to the ranges expected under reference conditions. Each metric that is outside of the reference range increases the IBI score by one. Therefore, if all four metrics were inside the reference range, the score would be 0. Conversely, if all four were outside the reference range, the value would be 4.

The data needed to calculate the IBI are the total number of taxa, number of mollusc taxa, abundance of *Notomastus* sp., and number of sensitive taxa (Table 1). The total number of taxa, number of mollusc taxa, and abundance of *Notomastus* sp. can be obtained directly from the data. The list of sensitive species should be based on the species list for Southern California Marine Bays and the percentage of sensitive taxa present is calculated as:

% sensitive taxa= (number of sensitive taxa/total number of taxa) x 100

The value for each metric is then compared to a reference range for that metric (Table 2). The IBI score is set to zero before comparison to the reference range. For each metric that is out of the reference range (above or below), the IBI score goes up by one.

Table 2. Reference ranges for IBI metrics in Southern California Marine Bays.

Metric	Reference Range
Total Number of Taxa	13 to 99
Number of Mollusc Taxa	2 to 25
Abundance of *Notomastus sp.*	0 to 59
Percentage of Sensitive Taxa	19 to 47.1

The IBI score is then compared to condition category thresholds (Table 3) in order to determine the IBI category and score.

Table 3. IBI category thresholds for Southern California Marine Bays.

IBI Score	Category	Category Score
0	Reference	1
1	Low Disturbance	2
2	Moderate Disturbance	3
3 or 4	High Disturbance	4

(ii) Relative Benthic Index (RBI) and RBI condition category

The RBI is the weighted sum of: (a) four community metrics related to biodiversity (total number of taxa, number of crustacean taxa, abundance of crustacean individuals, and number of mollusc taxa), (b) abundances of three positive indicator taxa, and (c) the presence of two negative indicator species.

The data needed to calculate the RBI are: total number of taxa, number of mollusc taxa, number of crustacean taxa, number of crustacean individuals, number of individuals of *Monocorophium insidiosum*, *Asthenothaerus diegensis*, and *Goniada littorea*, and the presence of *Capitella capitata* complex and Oligochaeta.

The first step is to normalize the values for the benthic community metrics relative to maxima for the data used to develop the RBI for the Southern California Marine Bays habitat, to produce values relative to the maxima that are referred to as scaled values. The scaled value calculations use the following formulae:

Total number of taxa / 99
Number of mollusc taxa / 28
Number of crustacean taxa / 29

The next step is to calculate the Taxa Richness Weighted Value (TWV) from the scaled values by the equation:

TWV = Scaled total number of taxa + Scaled number of mollusc taxa + Scaled number of crustacean taxa + (0.25 x Scaled abundance of Crustacea)

Next, the value for the two negative indicator taxa (NIT) is calculated. The two negative indicator taxa are *Capitella capitata* complex and Oligochaeta. For each of these taxa that are present, in any abundance whatsoever, the NIT is decreased by 0.1. Therefore, if neither were found the NIT=0, if both are found the NIT=-0.2.

The next step is to calculate the value for the three positive indicator taxa (PIT). The positive indicator taxa are *Monocorophium insidiosum*, *Asthenothaerus diegensis*, and *Goniada littorea.* First, the PIT value is calculated for each species using the following equations:

$$\frac{\sqrt[4]{Monocorophium\ insidiosum\ \text{abundance}}}{\sqrt[4]{473}}$$

$$\frac{\sqrt[4]{Asthenothaerus\ diegensis\ \text{abundance}}}{\sqrt[4]{27}}$$

$$\frac{\sqrt[4]{Goniada\ littorea\ \text{abundance}}}{\sqrt[4]{15}}$$

The three species PIT values are then summed to calculate the PIT value for the sample. If none of the three species is present, then the sample PIT = 0.

The next step is to calculate the Raw RBI:

Raw RBI = TWV + NIT + (2 x PIT)

The final calculation is for the RBI Score, normalizing the Raw RBI by the minimum and maximum Raw RBI values in the index development data:

RBI Score = (Raw RBI - 0.03)/4.69

The last step in the RBI process is to compare the RBI Score to a set of thresholds to determine the RBI category (Table 4).

5

Table 4. RBI category thresholds for Southern California Marine Bays.

RBI Score	Category	Category Score
> 0.27	Reference	1
> 0.16 to ≤ 0.27	Low Disturbance	2
> 0.08 to ≤ 0.16	Moderate Disturbance	3
≤ 0.08	High Disturbance	4

(iii) Benthic Response Index (BRI) and BRI condition category

The BRI is the abundance weighted pollution tolerance score of the organisms present in a benthic sample. The higher the BRI score, the more degraded the benthic community represented by the sample.

Two types of data are needed to calculate the BRI, the abundance of each species and its pollution tolerance score, P. P values are available for most species present in the assemblage. Only species for which P values are available are used in the BRI calculations. P values should be obtained for the appropriate habitat and from the most up-to-date list available.

The first step in the BRI calculation is to compute the 4th root of the abundance of each taxon in the sample for which P values are available. The next step is to multiply the 4th root abundance value by the P value, for each taxon.

Next, separately sum all of the 4th roots of the abundances and all of the products of the 4th roots of abundance and P values. Taxa that lack P values are not included in either sum.

The next step is to calculate the BRI score as:

$$\frac{\sum \left(\sqrt[4]{Abundance}\right) \times P}{\sum \sqrt[4]{Abundance}}$$

The last step is to compare the BRI score to BRI threshold values in Table 5 to determine the BRI category and category score.

Table 5. BRI category thresholds for Southern California Marine Bays.

BRI Score	Category	Category Score
< 39.96	Reference	1
≥ 39.96 to < 49.15	Low Disturbance	2
≥ 49.15 to < 73.27	Moderate Disturbance	3
≥ 73.27	High Disturbance	4

(iv) River Invertebrate Prediction and Classification System (RIVPACS) Index and RIVPACS condition category

The RIVPACS index calculates the number of reference taxa present in the test sample (observed or "O") and compares it to the number expected to be present ("E") in a reference sample from the same habitat. Calculation of the RIVPACS score is a three-step process. The first step consists of determining the probability of the test sample belonging to twelve Southern California Marine Bays reference sample groups. This determination is based on the sampling station's bottom depth, latitude, and longitude, using a complex linear discriminant function.

The second step is determining, for each sample, the identity and expected number of reference species, based on the probabilities of group membership calculated in Step 1 and the distribution of reference species in each group. In the final step, the number of reference species observed in the sample is counted, the O/E (Observed / Expected) RIVPACS score calculated and compared to the thresholds in Table 6 to determine the RIVPACS category and category score.

Table 6. RIVPACS category thresholds for Southern California Marine Bays.

RIVPACS Score	Category	Category Score
> 0.90 to < 1.10	Reference	1
> 0.74 to ≤ 0.90 or ≥ 1.10 to < 1.26	Low Disturbance	2
> 0.32 to ≤ 0.74 or ≥ 1.26	Moderate Disturbance	3
≤ 0.32	High Disturbance	4

Because of the complexity of the RIVPACS calculations, computer programs are used to determine the O/E values. Detailed instructions for calculating RIVPACS O/E values by two computer programs are provided in Appendices A and B. Appendix A contains

instructions for calculating RIVPACS O/E values using the Statistical Analysis System (SAS), while Appendix B contains instructions for calculating these values using the website at Utah State University's Western Center for Monitoring and Assessment of Freshwater Ecosystems. The SAS programs calculate RIVPACS O/E values and condition categories, but require availability of the SAS software. The Utah State University website is freely available to calculate RIVPACS O/E values, but data requirements are rigid and application of thresholds to determine condition categories is a separate procedure.

(v) Species list contents

The Southern California Marine Bays species list is provided as a spreadsheet with 12 columns in the **Embayment_Species_List_01_11_2008.xls** workbook. The contents of each column are described in Table 7.

Table 7. Southern California Marine Bays species list contents.

Column	Header	Contents
1	TaxonName	Taxon name
2	Phylum	Taxonomic phylum
3	Class	Taxonomic class
4	Order	Taxonomic order
5	Family	Taxonomic family
6	IBISensitive	When present, "S" indicates a taxon considered sensitive for calculation of the SoCal IBI
7	Mollusc	When present, "Mollusc" indicates molluscan taxa for RBI and IBI calculations
8	Crustacean	When present, "Crustacean" indicates crustacean taxa for RBI calculations
9	Tolerance Score	When present, values are tolerance scores for BRI calculation
10	RivColHead	When present, in the abundance data file submitted for RIVPACS calculations to the Utah State University web site, this exact text is used as the column header for abundance data for this taxon
11	RivColNo	When present, in the abundance data file submitted for RIVPACS calculations to the Utah State University web site, this is the column number containing abundances for this taxon.
12	SpeciesLevel	When present, "Drop" in this column indicates that abundances of this taxon are included in index calculations, but it is not included for counting numbers of taxa because lower taxonomic level entries in this taxon are also present.

B. Polyhaline Central San Francisco Bay

Table 8 presents details of the metrics included for calculation of index values in Polyhaline Central San Francisco Bay. Instructions for calculating each of the indices and descriptions of the species list variables follow.

Table 8. Benthic indicator metrics in Polyhaline Central San Francisco Bay. Metrics with superscript letters are included in both the RBI and IBI.

Index	Metric	Use
IBI	Total number of taxa[a]	All taxa
	Number of amphipod taxa	Amphipods
	Total abundance	All taxa
	Abundance of *Capitella capitata* complex	*Capitella capitata* Cmplx
RBI	Total number of taxa[a]	All taxa
	Number of mollusc taxa	Molluscs
	Number of crustacean taxa	Crustaceans
	Number of crustacean individuals	Crustaceans
	Abundance of *Sinocorophium heteroceratum*	*Sinocorophium heteroceratum*
	Abundance of *Rochefortia* spp.	*Rochefortia* spp.
	Abundance of *Prionospio (Minuspio) lighti*	*Prionospio (Minuspio) lighti*
	Presence of *Capitella capitata* complex	*Capitella capitata* Cmplx
	Presence of Oligochaeta	Oligochaeta
BRI	Abundance-weighted average tolerance score	ToleranceScore
RIVPACS	Observed to expected ratio for number of RIVPACS reference taxa.	Instructions for calculating O/E Ratio using SAS Software (Appendix A) or the Utah State University web site (Appendix B).

(i) Index of Biotic Integrity (IBI) and IBI condition category

The IBI compares the values of four different metrics to the ranges expected under reference conditions. Each metric that is outside of the reference range increases the IBI score by one. Therefore, if all four metrics were inside the reference range, the score would be 0. Conversely, if all four were outside the reference range, the value would be 4.

The data needed to calculate the IBI are the total number of taxa, number of amphipod taxa, total abundance, and abundance of *Capitella capitata* complex (Table 8), which can be obtained directly from the data.

The value for each metric is then compared to a reference range for that metric (Table 9). The IBI score is set to zero before comparison to the thresholds. For each metric that is out of the reference range (above or below), the IBI score goes up by one.

Table 9. Reference ranges for IBI metrics in Polyhaline Central San Francisco Bay.

Metric	Reference Range
Total Number of Taxa	21 to 66
Number of Amphipod Taxa	2 to 11
Total Abundance	97 to 2931
Abundance of *Capitella capitata* complex	0 to 13

The IBI score is then compared to condition category thresholds (Table 10) in order to determine the IBI category and score.

Table 10. IBI category thresholds for Polyhaline Central San Francisco Bay.

IBI Score	Category	Category Score
0 or 1	Reference	1
2	Low Disturbance	2
3	Moderate Disturbance	3
4	High Disturbance	4

(ii) Relative Benthic Index (RBI) and RBI condition category

The RBI is the weighted sum of: (a) four community metrics related to biodiversity (total number of taxa, number of crustacean taxa, abundance of crustacean individuals, and number of mollusc taxa), (b) abundances of three positive indicator taxa, and (c) the presence of two negative indicator species.

The data needed to calculate the RBI are: total number of taxa, number of mollusc taxa, number of crustacean taxa, number of crustacean individuals, number of individuals of *Sinocorophium heteroceratum*, the genus *Rochefortia*, and *Prionospio (Minuspio) lighti*, and the presence of *Capitella capitata* complex and Oligochaeta.

The first step is to normalize the values for the benthic community metrics relative to maxima for the data used to develop the RBI for the Polyhaline Central San Francisco Bay habitat, to produce values relative to the maxima that are referred to as scaled values. The scaled value calculations use the following formulae:

Total number of taxa / 55
Number of mollusc taxa / 13
Number of crustacean taxa / 17
Abundance of Crustacea / 17237

The next step is to calculate the Taxa Richness Weighted Value (TWV) from the scaled values by the equation:

TWV = Scaled total number of taxa + Scaled number of mollusc taxa + Scaled number of crustacean taxa + (0.25 x Scaled abundance of Crustacea)

Next, the value for the two negative indicator taxa (NIT) is calculated. The two negative indicator taxa are *Capitella capitata* complex and Oligochaeta. For each of these taxa that are present, in any abundance whatsoever, the NIT is decreased by 0.1. Therefore, if neither were found the NIT=0, if both are found the NIT=-0.2.

The next step is to calculate the value for the three positive indicator taxa (PIT). The positive indicator taxa are *Sinocorophium heteroceratum*, *Rochefortia* spp, and *Prionospio (Minuspio) lighti*. First, the PIT value is calculated for each species using the following equations:

$$\frac{\sqrt[4]{Sinocorophium\ heteroceratum\ abundance}}{\sqrt[4]{1878}}$$

$$\frac{\sqrt[4]{Rochefortia\ spp.\ abundance}}{\sqrt[4]{105}}$$

$$\frac{\sqrt[4]{Prionspio\ (Minuspio)\ lighti\ abundance}}{\sqrt[4]{17}}$$

The three species PIT values are then summed to calculate the PIT value for the sample. If none of the three species is present, then the sample PIT = 0.

The next step is to calculate the Raw RBI:

Raw RBI = TWV + NIT + (2 x PIT)

The final calculation is for the RBI Score, normalizing the Raw RBI by the minimum and maximum Raw RBI values in the index development data:

RBI Score = (Raw RBI - 0.00)/6.88

The last step in the RBI process is to compare the RBI Score to a set of thresholds to determine the RBI category (Table 11).

Table 11. RBI category thresholds for Polyhaline Central San Francisco Bay.

RBI Score	Category	Category Score
> 0.43	Reference	1
> 0.29 to ≤ 0.43	Low Disturbance	2
> 0.19 to ≤ 0.29	Moderate Disturbance	3
≤ 0.19	High Disturbance	4

(iii) Benthic Response Index (BRI) and BRI condition category

The BRI is the abundance weighted pollution tolerance score of the organisms present in a benthic sample. The higher the BRI score, the more degraded the benthic community represented by the sample.

Two types of data are needed to calculate the BRI, the abundance of each species and its pollution tolerance score, P. P values are available for most species present in the assemblage. Only species for which P values are available are used in the BRI calculations. P values should be obtained for the appropriate habitat and from the most up-to-date list available.

The first step in the BRI calculation is to compute the 4th root of the abundance of each taxon in the sample for which P values are available. The next step is to multiply the 4th root abundance value by the P value, for each taxon.

Next, separately sum all of the 4th roots of the abundances and all of the products of the 4th roots of abundance and P values. Taxa that lack P values are not included in either sum.

The next step is to calculate the BRI score as:

$$\frac{\sum \left(\sqrt[4]{Abundance}\right) \times P}{\sum \sqrt[4]{Abundance}}$$

The last step is to compare the BRI score to BRI threshold values in Table 12 to determine the BRI category and category score.

Table 12. BRI category thresholds for Polyhaline Central San Francisco Bay.

BRI Score	Category	Category Score
< 22.28	Reference	1
≥ 22.28 to <33.38	Low Disturbance	2
≥ 33.38 to <82.09	Moderate Disturbance	3
≥ 82.09	High Disturbance	4

(iv) River Invertebrate Prediction and Classification System (RIVPACS) Index and RIVPACS condition category

The RIVPACS index calculates the number of reference taxa present in the test sample (observed or "O") and compares it to the number expected to be present ("E") in a reference sample from the same habitat. Calculation of the RIVPACS score is a three-step process. The first step consists of determining the probability of the test sample belonging to four Polyhaline Central San Francisco Bay reference sample groups. This determination is based on the sampling station's bottom depth and longitude, using a complex linear discriminant function.

The second step is determining, for each sample, the identity and expected number of reference species, based on the probabilities of group membership calculated in Step 1 and the distribution of reference species in each group. In the final step, the number of reference species observed in the sample is counted, the O/E (Observed / Expected) RIVPACS score calculated and compared to the thresholds in Table 13 to determine the RIVPACS category and category score.

Table 13. RIVPACS category thresholds for Polyhaline Central San Francisco Bay.

RIVPACS Score	Category	Category Score
> 0.68 to < 1.32	Reference	1
> 0.32 to ≤ 0.68 or ≥ 1.32 to < 1.68	Low Disturbance	2
> 0.15 to ≤ 0.32 or ≥ 1.68	Moderate Disturbance	3
≤ 0.15	High Disturbance	4

Because of the complexity of the RIVPACS calculations, computer programs are used to determine the O/E values. Detailed instructions for calculating RIVPACS O/E values by two alternate computer programs are provided in Appendices A and B. Appendix A contains instructions for calculating RIVPACS O/E values using the Statistical Analysis

System (SAS), while Appendix B contains instructions for calculating these values using the website at Utah State University's Western Center for Monitoring and Assessment of Freshwater Ecosystems. The SAS programs calculate RIVPACS O/E values and condition categories, but require availability of the SAS software. The Utah State University website is freely available to calculate RIVPACS O/E values, but data requirements are rigid and application of thresholds to determine condition categories is a separate procedure.

(v) Species list contents

The Polyhaline Central San Francisco Bay species is list is provided as a spreadsheet with 12 columns in the **Embayment_Species_List_01_11_2008.xls** workbook. The contents of each column are described in Table 14.

Table 14. Polyhaline Central San Francisco Bay species list contents

Column	Header	Contents
1	TaxonName	Taxon name
2	Phylum	Taxonomic phylum
3	Class	Taxonomic class
4	Order	Taxonomic order
5	Family	Taxonomic family
6	Mollusc	When present, "Mollusc" indicates molluscan taxa for RBI calculations
7	Crustacean	When present, "Crustacean" indicates crustacean taxa for RBI calculations
8	Amphipod	When present, "Amphipod" indicates amphipod taxa for IBI calculations
9	Tolerance Score	When present, values are tolerance scores for BRI calculation
10	RivColHead	When present, in the abundance data file submitted for RIVPACS calculations to the Utah State University web site, this exact text is used as the column header for abundance data for this taxon
11	RivColNo	When present, in the abundance data file submitted for RIVPACS calculations to the Utah State University web site, this is the column number containing abundances for this taxon.
12	SpeciesLevel	When present, "Drop" in this column indicates that abundances of this taxon are included in index calculations, but it is not included for counting numbers of taxa because lower taxonomic level entries in this taxon are also present.

STEP 3: INTEGRATE BENTHIC INDEX CATEGORY SCORES

The final benthic community condition category is based on integrating all four benthic index category scores. The procedure is the same for samples from Southern California Marine Bays and samples from Polyhaline Central San Francisco Bay. Integration is accomplished by calculating the median of the four individual index category scores. If the median falls between two adjacent categories, the value is rounded up to the next highest integer category.

LITERATURE CITED

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. 2007. Evaluation of five indicators of benthic community condition in two California bay and estuary habitats. Technical Report 524. Southern California Coastal Water Research Project. Costa Mesa, CA.

APPENDIX A: USING THE STATISTICAL ANALYSIS SYSTEM (SAS) FOR RIVPACS CALCULATIONS

OVERVIEW

Two user-created files containing (1) habitat and (2) macrofauna data are submitted to a SAS program together with three provided data files containing (1) reference sample habitat data, (2) reference group species occurrence expectations, and (3) a master species list (Table A1). The SAS Program checks the samples in the test macrofauna data for presence in the habitat data, and taxon names against the master species list. Then it calculates RIVPACS O/E values and condition categories for each sample. Separate program and data files are provided for evaluating data from Southern California Marine Bays and Polyhaline Central San Francisco Bay.

METHODS

Running RIVPACS analysis using the Statistical Analysis System (SAS) computer program provided on the SCCWRP web site is a three-step process.
1. **Prepare data files.**
2. **Specify folder and file names.**
3. **Run the computer program.**

1. Prepare data files

The habitat data and macrofauna data for the test samples are prepared by the user as SAS data files, while the other necessary files (Table A1) are available on the SCCWRP web site. The user prepared files must contain one or more sample identifier variables; these variables are specified by the user. Examples of user prepared data files are provided on the SCCWRP web site. The example files use a single variable named "Site" to uniquely identify each sample.

In addition to identifier variables, habitat data files for Southern California Marine Bays contain SampleDepth (in meters) as well as Latitude and Longitude in decimal degrees (Table A2). Polyhaline Central San Francisco Bay habitat data files contain a Hab_G variable instead of latitude (Table A3). The Hab_G variable represents a dummy variable used in model development, and is zero for samples from Polyhaline Central San Francisco Bay. In all user created files, the data variable names and characteristics must exactly follow tables A2 to A4. The number of sample identifier variables and their names and characteristics are entirely at the user's discretion, but they must match exactly between the habitat and macrofauna data files.

Table A1. Computer files for calculating RIVPACS using SAS.

Contents	Source	File Name	
		Southern California Marine Bays	**Polyhaline Central San Francisco Bay**
SAS Program	Provided	RIVPACS_SouthernCaliforniaMarineBays.sas	RIVPACS_PolyhalineCentralSanFranciscoBay.sas
Master Species List	Provided	SasMasterBenthicTaxonListAll_20071231.sas7bdat	
Reference Habitat Data	Provided	SasSoCalMB_RefHabitat.sas7bdat	SasPCSFB_RefHabitat.sas7bdat
Reference Species List	Provided	SasSoCalMB_RefSpecies.sas7bdat	SasPCSFB_RefSpecies.sas7bdat
Reference Species Expectations	Provided	SasSoCalMB_RefSpeciesExpect.sas7bdat	SasPCSFB_RefSpeciesExpect.sas7bdat
Test Sample Habitat Data	User Created	SasSoCalMB_TestHabitatExample.sas7bdat	SasPCSFB_TestHabitatExample.sas7bdat
Test Sample Macrofauna Data	User Created	SasSoCalMB_TestFaunaExample.sas7bdat	SasPCSFB_TestFaunaExample.sas7bdat

Table A2. Contents of user created habitat data files for Southern California Marine Bays.

Variable	Type	Length	Format	Units
Latitude	Numeric	8 bytes	11.6	Decimal degrees
Longitude	Numeric	8 bytes	11.6	Decimal degrees
SampleDepth	Numeric	8 bytes	4.1	Meters
Sample Identifier 1	User's choice, but consistent with macrofauna data			
Sample Identifier 2	User's choice, but consistent with macrofauna data			
Sample Identifier n	User's choice, but consistent with macrofauna data			

Table A3. Contents of user created habitat data files for Polyhaline Central San Francisco Bay.

Variable	Type	Length	Format	Units
Hab_G	Numeric	8 bytes	1.0	None
Longitude	Numeric	8 bytes	11.6	Decimal degrees
SampleDepth	Numeric	8 bytes	4.1	Meters
Sample Identifier 1	User's choice, but consistent with macrofauna data			
Sample Identifier 2	User's choice, but consistent with macrofauna data			
Sample Identifier n	User's choice, but consistent with macrofauna data			

For both habitats, the test macrofauna data is a simple file with taxon names and abundances on each row in addition to the identifier variable(s). Each row contains the name and abundance for one taxon (Table A4).

Table A4. Contents of user created macrofauna data files for Southern California Marine Bays and Polyhaline Central San Francisco Bay.

Variable	Type	Length
Abundance	Numeric	8 bytes
TaxonName	Character	50
Sample Identifier 1	User's choice, but consistent with habitat data	
Sample Identifier 2	User's choice, but consistent with habitat data	
Sample Identifier n	User's choice, but consistent with habitat data	

2. Specify folder, identifier variable, and file names

Edit the five lines that immediately follow the instructions at the beginning of the program file:

- Specify the folder where the data are stored in the Libname statement;
- Specify the names of the variables in your data that uniquely identify each sample in the "identifier" instruction. The example data contain a single identifier variable named "Site."

- Specify the names of the files containing habitat and macrofauna data to be analyzed, and the name of a file to be created to store results.

3. Run the computer program

Submit the data to run in SAS. The program output includes the identifier variables, RIVPACS O, E, and O/E values, and condition categories for each sample. Two types of output are generated: an output window and a stored results file. The output window contains the analysis results and describes the contents of the SAS data file that is created to store them. The stored results file can be used to facilitate data transfer or subsequent data analysis.

APPENDIX B: USING THE WEB SITE AT UTAH STATE UNIVERSITY'S CENTER FOR MONITORING AND ASSESSMENT OF FRESHWATER ECOSYSTEMS FOR RIVPACS CALCULATIONS

OVERVIEW

Two files containing (1) habitat and (2) macrofauna data are prepared by the user and submitted to the web site at Utah State University's Center for Monitoring and Assessment of Freshwater Ecosystems together with three provided data files containing (1) reference sample group means, (2) reference sample inverse covariance matrix, and (3) reference sample macrofauna data (Table B1). The web site calculates RIVPACS O/E values and the results can be saved on the user's computer. Different reference data files are used for data from Southern California Marine Bays and Polyhaline Central San Francisco Bay (Table B1).

METHODS

Running RIVPACS analysis on the web site at Utah State University's Center for Monitoring and Assessment of Freshwater Ecosystems is a four-step process:

1. **Get a Username and Password.**
2. **Prepare data files.**
3. **Login and upload data files.**
4. **Save results locally and compare O/E values to condition category thresholds.**

1. **Get a Username and Password**

   Send an e-mail requesting a user name and password to Dr. Chuck Hawkins [chuck.hawkins@usu.edu]. In the e-mail, mention that you intend to run the predictive model prepared for the California Sediment Quality Objectives program. Allow a few days for a response.

Table B1. Computer files for calculating RIVPACS using the web site at Utah State University's Center for Monitoring and Assessment of Freshwater Ecosystems.

Contents	Source	Web Site File Reference	File Name	
			Southern California Marine Bays	**Polyhaline Central San Francisco Bay**
Reference Habitat Group Means	Provided	Groupmeans	SoCalMB_Ref_GroupMeans.txt	PCSFB_Ref_GroupMeans.txt
Reference Habitat Inverse Covariance Matrix	Provided	Inverse covariance values	SoCalMB_Ref_InvCovariance.txt	PCSFB_Ref_InvCovariance.txt
Occurrence of Macrofauna Species in Reference Groups	Provided	Reference bugs	SoCalMB_Ref_Fauna.txt	PCSFB_Ref_Fauna.txt
Test Sample Habitat Data	User Created	Test habitat	SoCalMB_Test_Habitat_Example.txt (Example)	PCSFB_Test_Habitat_Example.txt (Example)
Test Sample Macrofauna Data	User Created	Test bugs	SoCalMB_Test_Fauna_Example.txt (Example)	PCSFB_Test_Fauna_Example.txt (Example)

2. **Prepare data files**

Download the appropriate files from the SCCWRP web site into a folder where the files for this analysis are to be stored. Five data files are required to run the model in each of the habitats for which models have been developed. Although the types of files are common to the Southern California Marine Bays and Polyhaline Central San Francisco Bay, details of the file formats vary slightly between the habitats.

Three of the files contain reference site information used to create the model and are distributed by SCCWRP, including: the "groupmeans" file, the "inverse covariance values" file and the "reference bugs" file. Different sets of reference files are provided for data from Southern California Marine Bays and Polyhaline Central San Francisco Bay (Table B1). The other two files contain habitat data and macrofauna data for the test sites for which RIVPACS O/E values are to be calculated. The web site software is very sensitive to file format and returns error messages instead of results unless the habitat and macrofauna data files are formatted perfectly. The habitat and macrofauna data files are both tab-delimited with a single row for each sample. The first column in the habitat data and macrofauna data files contains a sample identifier that includes no spaces, and the samples are in the same order for both files. The habitat data files contain the information included in Table B2.

The macrofauna data files contain sample identifiers in the first column and species abundances for reference species in subsequent columns. Macrofauna data files for Southern California Marine Bays include abundance data for 457 reference taxa, while macrofauna data files for Polyhaline Central San Francisco Bay include data for 119 reference taxa. Zero abundance is represented by zero. The web software produces an error message if blanks represent zero abundance. The web software is sensitive to the order and exact spelling of the column headers with taxon information. The header names and order for each habitat are provided in the species list for each habitat. The header is the first 32 characters of the taxon name with blanks and parentheses replaced by underscores (_). Trailing blanks are not replaced.

Example files that include habitat and macrofauna data in correct formats for 15 samples each from Southern California Marine Bays and Polyhaline San Francisco Bay are provided along with the reference files on the SCCWRP web site. It is convenient to keep all the files needed for a particular analysis together in a single folder.

Table B2. Contents of habitat data files for web submission.

Column	Header	Contents
		Southern California Marine Bays
1	Any	Sample identifier
2	Latitude	Latitude in decimal degrees
3	Longitude	Longitude in decimal degrees
4	SampleDepth	Bottom depth in meters
		Polyhaline Central San Francisco Bay
1	Any	Sample identifier
2	SampleDepth	Bottom depth in meters
3	Hab_G	Zero for samples in Polyhaline Central San Francisco Bay
4	Longitude	Longitude in decimal degrees

3. **Login and upload data files**

    Until you receive a username and password, you can browse general information about predictive modeling and do everything other than run the models at: http://129.123.10.240/WMCPortal/DesktopDefault.aspx?tabindex=0&tabid=1.

    Login using the username and password provided by Chuck Hawkins to the web site at: http://wmc2.bnr.usu.edu:8080/examples/servlets/LoginSession.html. Enter the username and password in the appropriate spaces, and click on "Login."

    On the next page, use the defaults of "New model" for "Name of the model" and "Tab delimited" for "Delimiter" and click on "Continue." On the next page, identify the five requested files. Pick the "Browse" option for the first file, locate it on your computer, and click on it. The website software is quite smart and will look first in the same folder for the next file. The files to be identified (Table B1) are: "groupmeans", "inverse covariance values", "reference bugs", "test bugs", and "test habitat".

    Once all five files are identified, click on "Submit data." If the model runs successfully, an "Output files" page containing four sets of results will appear. If the run is unsuccessful, the test data file formats do not meet the expectations of the software and an error message will display. Correct the format error and resubmit the files.

4. **Save results locally.**

    Links to four sets of results files are present on the "Output files" page (Table B3). Each set is available in HTML format and text format. Files in HTML format are easiest to view and are easily copied and pasted into spreadsheets. Results files can also be saved on the user's computer by right-clicking on the link

to the results on the "Output files" page and selecting "Save Target As." Note that this instruction applies to Internet Explorer; the method used to save the results locally may vary for other internet browsers.

The "O over E scores" results contain the only information required for comparison with the RIVPACS category thresholds. Only the O/E scores for P > 0.5 contained in the column to the far right are used for evaluating sample condition and assigning condition categories. The P > 0.5 O/E scores are compared to thresholds in Table 6 in the main body of this document for samples in Southern California Marine Bays or Table 13 for samples in Polyhaline Central San Francisco Bay.

The "O over E scores" results file contains O (Observed), E (Expected), and O/E (Observed over Expected) ratios for reference species at two probability of detection thresholds: P > 0.0 and P > 0.5. The higher probability threshold is used for assessment purposes because rare taxa cannot be modeled as well as more common taxa, and errors in their prediction lead to errors in E. In general, models based on intermediate probability of detection thresholds (e.g., P > 0.5) yield models that are more precise and more sensitive in detecting effects of stressors.

Information about the other three types of results is presented in Table B3. Additional information on subjects related to RIVPACS is available from Utah State University's web site at:
http://129.123.10.240/WMCPortal/DesktopDefault.aspx?tabindex=0&tabid=1.

5. **When done, click "Finish" and log out**

Table B3. Results produced on the "Output files" page at the web site at Utah State University's Center for Monitoring and Assessment of Freshwater Ecosystems.

Results Type	Contents	Use for Categorical Assessment
O over E scores	Observed (O) and expected (E) numbers of species and O/E ratios for reference species with probabilities of occurrence > 0.0, and > 0.5.	Thresholds applied to O/E ratios for P > 0.5.
Site Habitat Test Results	A summary of site data, including habitat variables and total organism counts. A "P" value in the Model Test column indicates that values of the predictor variables measured at the sites being assessed are within a statistically acceptable range of values measured at the reference sites.	Not used. But if the Model Test column is not "P" ("F" or "Fail"), the test site has values that fall outside the predictor reference range, and the user should evaluate whether the extrapolation is valid or not.
Probability of taxa occurrences at sites	The probability matrix indicates the probability of occurrence for each taxon at each site. Taxa expected to occur (p > 0.5) that were not present are highlighted in red. Taxa that were not expected to occur (p < 0.5) but were present are highlighted in green.	Not used. Supplementary information for interpretive purposes.
Taxon occurrence summary	Summarizes taxon expectations and occurrences. The sensitivity index is an O/E ratio for the taxon where 0 indicates absence from test sites, 1 indicates occurrence at expected frequency, and values > 1.0 indicate occurrence more often than expected.	Not used. Supplementary information for interpretive purposes.