

## **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.

#### 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)

### 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
<b>Southern California Marine Bays</b>		
1	Any	Sample identifier
2	Latitude	Latitude in decimal degrees
3	Longitude	Longitude in decimal degrees
4	SampleDepth	Bottom depth in meters
<b>Polyhaline Central San Francisco Bay</b>		
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

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.

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 5.8 in the main body of this document for

samples in Southern California Marine Bays or Table 15 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>.

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.**

<b>Results Type</b>	<b>Contents</b>	<b>Use for Categorical Assessment</b>
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.