

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.

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 Sample Depth (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	

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.

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.