Ecosons Split-beam analysis: software manual

Daniel Rodríguez-Pérez
Noela Sánchez-Carnero

28 de octubre de 2011

Ecosons is a software for acoustic data processing. Among its options is texture calculation from angular data from split-beam echosounder files. The following manual explains how to perform the textural analysis of angular data using the program "ecosons splittex" for Octave downloadable from

`http://www.recursosmarinos.net/downloads/haralick`.

Once the file ecosons_splittex.zip is decompressed in the destination folder (selected by the user), the octave program is run in this folder. Under Microsoft Windows, the octave program will be run from a shortcut in the "Start menu" (created by the Windows installer that can be downloaded from `http://octave.sourceforge.net/`); once in the octave console, the user will have to change to the folder where ecosons_splittex.zip had been decompressed ("c:\splitbeam", for instance) using the instruction:

```
cd 'c:/splitbeam'
```

The split-beam texture analysis program is started with the "splittex" command, opening a 6-option menu:

- Split beam angles texture processor:

- [1] Load split beam transect data

- [2] Load transect data

- [3] Input algorithm parameters

- [4] Run processor

- [5] Export features

- [6] Display classification results

- [7] Quit

Figura 1: Acoustic data organization and classification file.

[1] Load split beam transect data

First, the user has to load a file informing the program of the organization and classification of the acoustic files that will be subsequently processed; this file must be located in the same folder as the acoustic files.

- Survey structure file (default survey.dat): split_data/survey.dat

This file must be a plain text file containing, for each transect, its label, orientation and the names of the corresponding acoustic files (without the extension). Each of these data must be in a row, as shown in figure 1. Consecutive transects must be consecutively stored in this file, separated by a blank line. Lines starting with "#" will be ignored, which can be useful for adding comments or excluding files from the study without the need to delete them.

[2] Load transect data

Once the user has loaded the survey organization file, acoustic files are read. All files must be in the same folder as the survey file ("survey.dat").

In order to speed up the reading and processing of acoustic data, the user must define the region of interest (ROI) in meters, defined as the echogram portion that will be studied below bottom detection (30 cm by default).

- Input ROI depth (default 0.30 m):

While the files are processed, the software informs about the processing progress.

- Rb1 Converting datos_split/bu-D20090612-T095413.raw.

- Read ping no. 1

- Read ping no. 2

- ...

[3] Input algorithm parameters

In this section, the user must define the parameters for the analysis that will be done in the next steps. This application is based on the calculation and statistical analysis of textural variables (defined by Haralick, 1973) which are calculated on a co-occurrence matrix built from angular echograms (longitudinal and transversal angle) from the acoustic files of a split-beam echosounder.

The user must select the parameters required to define the process of co-occurrence matrix definition and the variables computation.

As a first step for matrix calculation, angular matrices must be normalized, coding their values as a function of their probability distribution (mean, μ and standard deviation, σ). This probability distribution can be computed bin by bin, along each ping [2] or bin by bin at a fixed depth (going from one ping to the next) [1]. As a third option, the user may code the matrix without using μ and σ, by establishing interval coding in degrees [3].

- Select data normalization direction:

- [1] Across pings (along constant depth)

- [2] Along pings

- [3] No normalization

Once the method for calculating the mean and standard deviation of the sample are defined, the user must define the codification to be used. By default, a codification of 4 values is proposed:

$$
\begin{array}{ll}
1, & \text{si } \mu - 2\sigma < x < \mu - \sigma \\
2, & \text{si } \mu - \sigma < x < \mu \\
3, & \text{si } \mu < x < \mu + \sigma \\
4, & \text{si } \mu + \sigma < x < \mu + 2\sigma
\end{array}
$$

- Input quantization levels in sigma units (default [-2 -1 0 1 2]):

- 4 quantization intervals selected.

When the angular matrix has been coded, the co-occurrence matrix is generated through a bin-by-bin comparison of the values in the coded matrix. This comparison can be done, as in the previous step, bin by bin through each ping [2], bin by bin at a fixed step (going from one ping to the next) [1] or bin by bin in both directions [3].

- Select texture direction:

- [1] Across pings

- [2] Along pings
- [3] First neighbors

The user can define the co-occurrence matrix computation method, i.e. if the bin-by-bin comparison will be done along adjoining bins (default option) or if it will be performed between bins separated by x positions (in the selected working direction).

- Select texture step length (default, 1):

The user has also the option to subdivide each transect into segments with an equal number of pings. If this option is not desired, a single segment can be chosen, which will process each transect as a study unit.

- Choose number of segments per transect (default 4):

Finally, the user can choose between calculating the textural variables defined by Haralick (1973) or obtaining the coefficient values directly from the co-occurrence matrix for subsequent analyses.

- Select features to compute:
- [1] Haralick textural features
- [2] Cooccurrence matrix coefficients

[4] Run processor

With this option, the user can start the process as defined in the previous option ([3] Input algorithm parameters).

- Processing Rb1
- Processing Re2
- Processing Rb3
- ...

[5] Export features

This option allows the user to export data matrices generated in the previous option ([4] Run processor).

- Input output file name (default ./splittex.dat):

The structure of the text file generated is the following: columns contain the transect number, the description of its orientation regarding the coastline, the segment number, the starting ping and the final ping of each transect (considering the different acoustic files which constitute it), the latitude and longitude coordinates of the center of the segment, and the calculated characteristics, first for the Al (along-ship) angles, and then for the At (athwart-ship) angles. Rows vary segment and transect number. The first row in the file contains the names of the different columns (headers).

[6] Display classification results

This option allows the user to perform a statistical classification from the variables generated in the previous steps. The analysis may be done with all the data or include a subgroup of transects (according to their orientation, or previously defined in the survey file): portboard, starboard.

- Filter by description ("" none): starboard

First, the user can perform a principal component analysis (PCA) and reduce the dimensionality of the matrix.

- Apply PCA to the normalized feature matrix? (Y/n)
- Save PCA coefficients? (blank to skip)

The program returns the number of principal components (PC) selected using Kaiser's rule, and the 5 most representative variables within the two first components.

- Selected 11 PCs with eigenvalues larger than 1
- The 5 most relevant features in the first 2 components are:
- [1] 24
- [2] 12
- [3] 26
- [4] 15
- [5] 19
- Best silouhette number: 0.36738 for 3 (<=10) classes

The program also allows to perform a cluster analysis (with Euclidean distances and grouping by average method). The default number of classes is suggested based on the "silhouette number", which indicates the best separation into different classes, but the user can chose a different number of classes.
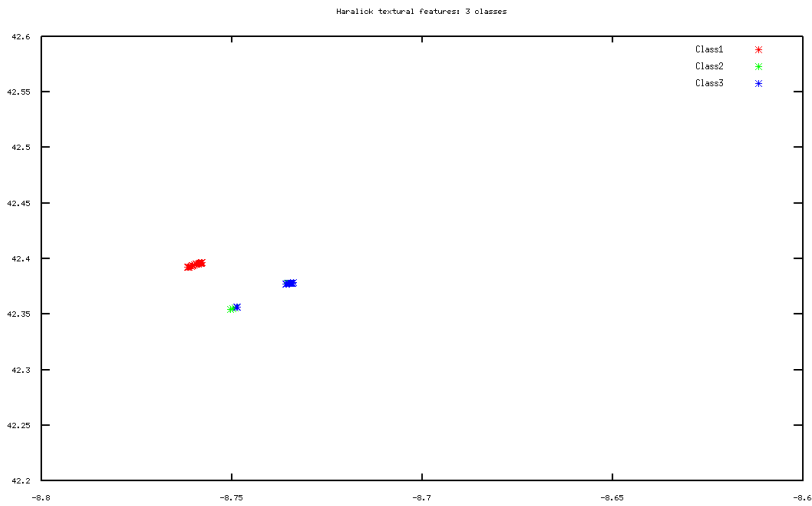
Figura 2: Plot showing transect centroids classified in three classes.

- Select number of classes (default 3): 3

- Silouhette number: 0.36738

  CmxL =

  Columns 1 through 31:

  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 21 21 21 21 25 25 25 25 21 21 21

  Columns 32 through 56:

  21 25 25 25 25 21 21 21 21 21 25 25 25 25 25 1 21 25 25 1 21 21 25 25 25

The program shows the vectors with the numerical labels assigned to the consecutive segments of the transects, as well as the result of the classification, in a plot (figure 2). This plot can be saved as a .fig file (editable with http://www.xfig.org/ or with http://tams-www.informatik.uni-hamburg.de/applets/jfig/).

- Save as FIG file (blank to skip):