Automatic seabed classification using functional data analysis and time series cluster techniques

Andrés Prieto[1], Noela Sánchez-Carnero[2], Javier Tarrío-Saavedra[3,*]

[1] *Grupo M²NICA. Facultade de Informática. Universidade da Coruña. Spain; andres.prieto@udc.es*
[2] *Grupo de Oceanografía Física. Universidade de Vigo. Spain; noelas@gmail.com*
[3] *Grupo MODES. Universidade da Coruña. Escola Politécnica Superior. Spain; javier.tarrio@udc.es*
[*] *Corresponding author*

Abstract. *Seabed characterization in coastal environments is usually based on acoustic techniques. Since intrusive measurements are very time-consuming, data acquired by echosounders are the best option for classification purposes. The acoustic seabed response is measured by recording local averages of the intensity field during a time interval, which contains the first echo produced by a sonar pulse excitation emitted from the water surface. The standard methodology for the sea bottom classification relies on the accurate extraction of features, which enable a classical multivariate cluster analysis. The effectivity of such reduction of dimensionality on the data may be enhanced by a preprocessing of the signals based on physical knowledge about the acoustic behaviour of the intensity curves depending in the relative position of the echosounder with respect to the seabed. The automatic seabed classification proposed in this work is performed by means of either time series cluster methods or functional data analysis (FDA) non-hierarchical cluster techniques. In both cases, this method does not require any a priori knowledge of the feature extraction on the sonar curves. More precisely, unsupervised methods such as the FDA K-means method, the multivariate medoids cluster, and time series cluster techniques have been applied. The supervised FDA techniques such as functional generalized linear models (GLM) and generalized spectral additive models (GSAM) have been also considered. The proposed technique is illustrated with some sonar data measured in a controlled environment (where the real classification is well-known) and compared with those results obtained with classical multivariate hierarchical cluster tools.*

Keywords. *Cluster classification; Functional Data Analysis; Time Series.*

---

1 Introduction

Acoustic data obtained with sonar devices provide important information about seabed characteristics: bathymetry, bottom type, presence or absence of vegetation, etc. Therefore, this type of data have been widely used for benthic habitat mapping [1, 9]. The analysis of these data has been addressed from different perspectives (multivariate statistics, spectral analysis, energy integration, etc). Multivariate methods of supervised and unsupervised seabed classification are usually applied to databases composed by relevant features extracted from acoustic echoes. The selection of representative features is by no means trivial. Currently, there are several alternatives that combine feature selection, taking into account the physical meaning of the acoustic response, with statistical dimensional reduction procedures, such as

Principal Component Analysis (PCA) [6, 10]. However, the automatic classification of seabed is still an open problem. The application of classification techniques involving Functional Data Analysis (FDA) [11] and also time series classification (based on dissimilarity measures) are promising approaches on this field. The main advantage of FDA and time series methods lies in the fact that their uses do not require an *a priori* knowledge of the physical problem, what is absolutely necessary to extract accurately the vector of features from the intensity curves, a must in most standard classification methods.

2 Data description and acoustic preprocessing

The acoustic survey was carried out on July 2014 with excellent weather conditions (without wind or waves on the water surface) from a small boat (~ 5 m length). Acoustic data were recorded with a single-beam echosounder (EA400P Simrad), working with a 38 and 200 kHz transducer, at three different sampling points (two over sandy bottoms: fine sand (class CP01) and fine sand with vegetation (CP02) and one over rocky bottom (CP03)) located near the Marine Protected Area of Cabo de Palos (Murcia, Spain). Overall 1383 curves have been obtained: 941 for CP01 class, 160 for CP02, and 282 corresponding to CP03 (see Figure 1). At each survey point several recordings were done varying pulse length, power and depth. This last factor was modified by lowering down the transducer vertically using a rope.

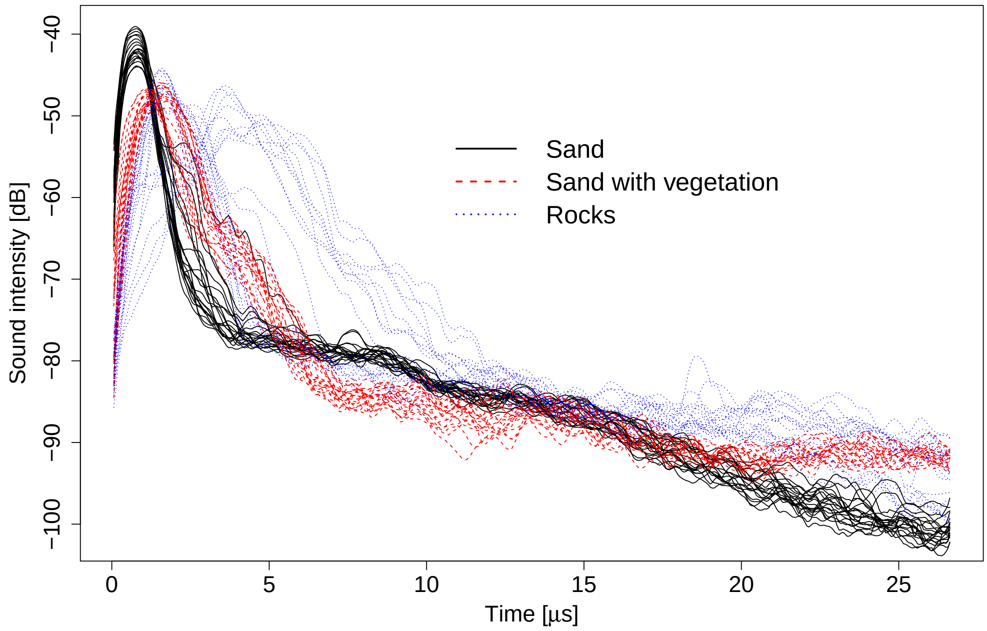

Figure 1: First echo recorded at 200 kHz, plotted for some of the corrected intensity curves of each seabed class. Black solid lines: CP01 class, red dashed lines: CP02 class and blue dotted lines: CP03 class.

The dataset consists in all the echoes recorded, allowed us to analyse the effect of different settings on the bottom acoustic response. In this work only data with the same pulse length were analysed (256 μs for 200 kHz and and 10246 μs for 38 kHz), since previous analyses showed that changes in pulse length introduce relevant modifications in echo shape that would need a special correction. Acquired acoustic data have to be corrected in order to remove the spreading and attenuation effects. In this sense, after bottom detection, time and power adjustment have been applied to raw data (see Figure 1). Moreover, corrected echoes were convoluted with a stretching kernel that made echo lengths comparable [13]. All these corrections were computed using ECOSONS software [14].

3 Supervised and cluster procedures

The present work proposes automatic seabed classification by means of supervised and unsupervised methods. For supervised classification, under a FDA perspective, generalized linear models (GLM) and generalized spectral additive models (GSAM) have been applied to estimate the seabed class (used as response variable) depending on acoustic intensity curves (considered as functional independent variable). Both functional variables and functional parameters have been appropriately represented in a B-spline basis [11].

Unsupervised classification has been performed using time series cluster procedures and FDA techniques. In this last case, a FDA version of the K-means algorithm has been applied [3]. Concerning to the time series classification, different measures of dissimilarity between time series have been computed before cluster algorithms application based on different approaches [7]. To evaluate the performance of the unsupervised classification methods, three different indicators have been considered: i) Cohen's κ index, ii) the classification accuracy (both computed from the contingency table values), and iii) the Clustering Evaluation Index (CEI) based on known ground truth [4], which computes the similarity between the right known classification and that one obtained with a method under evaluation.

Supervised and unsupervised methodologies have been applied to seabed echoes data and compared with the classical multivariate approach using hierarchical clustering (using Ward's linkage method and the Euclidean distance). R libraries `fda.usc`, `fpc`, `TSclust`, and `stats` have been used to perform the classification tasks. Although the dataset has been acquired at just three sampling points (thus minimizing variability among intensity curves within the same group), the computed classification has been significantly better than those obtained in [8, 15] (both of them using QTC software, the *de facto* standard for single-beam seabed classification) and even than those recently obtained in [2, 5, 12], where multibeam data is analysed (the gold standard echosounder with highest acoustic resolution). Moreover, in the present dataset two bottom types are remarkably similar (both classified *in situ* as fine sand) with only differences in roughness and compacity levels, illustrating that FDA and time series methodologies represent a promising new approach for statistical bottom echo classification.

In conclusion, the application of FDA and time series cluster techniques provides classification estimates highly concordant with the real classes of seabed (from supervised and unsupervised approaches). The results are highly competitive with those obtained by existing classical multivariate methods based on features extraction with a required *a priori* expertise. In fact, the proposed FDA and time series classification methodologies prevents the previous features selection and allow to automatically classify seabed without necessity of an expert knowledge on the echo curves.

Acknowledgments. This work has been supported by Xunta de Galicia project "Numerical simulation of high-frequency hydro-acoustic problems in coastal environments - SIMNUMAR" (EM2013/052), cofunded with European Regional Development Funds (ERDF).

References

[1] Brown, C. J., Smith, S. J., Lawton, P., Anderson, J. T. (2011). Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuarine, Coastal and Shelf Science*, **92**(3), 502–520.

[2] Elvenes, S., Dolan, M. F. J., Buhl-Mortensen, P., Bellec, V. K. (2014). An evaluation of compiled single-beam bathymetry data as a basis for regional sediment and biotope mapping. *ICES Journal of Marine Science: Journal du Conseil*, **71**(4), 867–881.

[3] Febrero-Bande, M., Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: The R package fda.usc. Journal of Statistical Software, **51**(4), 128.

[4] Gavrilov, M., Anguelov, D., Indyk P., Motwani, R. (2000). Mining the stock market (extended abstract): which measure is best? In Proceedings of the sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 487–496.

[5] Hasan, R. C., Ierodiaconou, D., Laurenson, L. (2012). Combining angular response classification and backscatter imagery segmentation for benthic biological habitat mapping. *Estuarine, coastal and shelf science*, **97**, 1–9.

[6] Legendre, P., Ellingsen, K. E., Bjørnbom, E., Casgrain, P. (2002). Rapid communication/communication rapide acoustic seabed classification: improved statistical method. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**(7), 1085–1089.

[7] Montero, P., Vilar, J. A. (2014). TSclust: An R Package for Time Series Clustering. Journal of Statistical Software, **62**(1), 1–43.

[8] Moyer, R. P., Riegl, B., Banks, K., Dodge, R. E. (2005). Assessing the accuracy of acoustic seabed classification for mapping coral reef environments in south florida (Broward County, USA). *Revista de Biología Tropical*, **53**, 175–184.

[9] Pandian, P. K., Ruscoe, J. P., Shields, M., Side, J. C., Harris, R. E., Kerr, S. A., Bullen, C. R. (2009). Seabed habitat mapping techniques: an overview of the performance of various systems. *Mediterranean Marine Science*, **10**(2), 29–44.

[10] Preston, J. M., Kirlin, T. L. (2003). Comment on "Acoustic seabed classification: improved statistical method", Canadian Journal of Fisheries and Aquatic Sciences, **60**, 1299–1300.

[11] Ramsay, J. O., Silverman, B. W. (2005). Functional Data Analysis. 2nd ed. Springer-Verlag. New York.

[12] Rattray, A., Ierodiaconou, D., Laurenson, L., Burq, S., Reston, M. (2009). Hydro-acoustic remote sensing of benthic biological communities on the shallow south east australian continental shelf. *Estuarine, Coastal and Shelf Science*, **84**(2), 237–245.

[13] Rodríguez-Pérez, D., Sánchez-Carnero, N., Freire, J. (2014). A pulse-length correction to improve energy-based seabed classification in coastal areas. *Continental Shelf Research*, **77**, 1–13.

[14] Rodríguez-Pérez, D., Sánchez-Carnero, N., Freire, J. (2014). ECOSONS software. (available at http://www.kartenn.es).

[15] Schimel, A. C. G., Healy, T. R., Johnson, D., Immenga, D. (2010). Quantitative experimental comparison of single-beam, sidescan, and multibeam benthic habitat maps. *ICES Journal of Marine Science: Journal du Conseil*, page fsq102.