{rokbox title=|Original echograms from the SCAPA survey at three frequencies after denoising with the algorithm in Peña, 2016 :: Figure: Authors| thumb=|images/stories/ieo/imagenespublicaciones/centro-oceanografico-baleares-robust-clustering-methodology-multi-frequency-acoustic-data-pena-2018-thumb.jpg|}images/stories/ieo/imagenespublicaciones/centro-oceanografico-baleares-robust-clustering-methodology-multi-frequency-acoustic-data-pena-2018.jpg{/rokbox}

Marian Peña, 2018. **Robust clustering methodology for multi-frequency acoustic data: A review of standardization, initialization and cluster geometry.**
Fisheries Research, 200 (2018). 49–60.

Abstract: Clustering is a useful unsupervised technique for the identification of acoustic groups in multi-frequency echograms based on frequency response. K-Means is the most well-known clustering technique but has significant requirements such as clusters of equal size and spherical shape. Initialization is a common problem in clustering as only local minima are usually guaranteed, and thus initialization must locate the centroids near the global minimum. Expectation-Maximization (EM) clustering also requires a good set of initial centroids but allows the identification of clusters with different statistical distributions. This work presents the comparison of these techniques applied to a case with several acoustic signatures presenting different cluster sizes and distributions. The main issues treated in this manuscript are: pre-processing of acoustic data for clustering, initialization of centroids with theoretical scattering models and the need to consider the geometry of the clusters in addition to means, including variance (spread around the mean), orientation (correlation between variables), spherical or ellipsoidal shape (difference in variance between variables) and cluster size (number of observations). EM clustering is the only technique that properly separates acoustic signatures (and noise) after using the supervised initialization presented in this study.

Keywords: Cluster geometry, Correlation, Frequency response, Standardization, Variance, K-Means, EM clustering