Supplement of Ocean Sci., 17, 1563–1583, 2021
https://doi.org/10.5194/os-17-1563-2021-supplement

Ocean Science Open Access EGU

Supplement of

Background stratification impacts on internal tide generation and abyssal propagation in the western equatorial Atlantic and the Bay of Biscay

Simon Barbot et al.

Correspondence to: Simon Barbot (simon.barbot@legos.obs-mip.fr)

Supplementary Material A: Tests to set the best parameters and methods for the clustering

The clustering methods can classify the different profiles in several clusters using a matrix of the distances from every profile to the other ones. Performing a principal component analysis (PCA) with two principal components on the profiles enables us to transform a system of $(N \times D)$ (with N the number of profiles and D the number of depths) to a system of $(N \times 2)$. In such a system, named the PCA manifold, the distance matrix is easily calculable (Fig. 1).

Three different methods of clustering have been tested: Ward, Average and Spectral (Python SciKitLearn clustering package[1]; Pedregosa et al., 2011). Those methods have been selected because they can better classify the distribution of the PCA manifold (Fig. 1). The Average hierarchical clustering method builds the complete tree that links the points by minimizing the average of the distances between the clusters being merged in order to build the tree (WPGMA, Sokal, 1958). The Ward hierarchical clustering method is based on the same methodology as Average but minimizes the variance of the clusters being merged (Ward, 1961; Ward and Hook, 1963). The Spectral clustering method is different from the previous two. This method projects the PCA manifold onto a polar coordinate space before performing the classification through a specified number of clusters and minimizing the distance within each cluster (Yu and Shi, 2003; Von Luxburg, 2007).

For these three methods, the sensitivity of two parameters needs to be investigated: the number of final clusters and the number of neighbors used in the calculation of the distance between profiles. The number of neighbors is important to properly manage the profiles that are isolated outside the PCA manifold. If the number of neighbors is weaker than the number of outsider profiles, then they all will be grouped in a dedicated cluster. Otherwise, they will be included in the cluster of the nearest profiles. This latter case can lead to groups of profiles that do not have the same shape inside the same cluster. The number of neighbors also affects the profiles located at the boundary between two clusters: depending on the number of neighbors, they will be included in one cluster or another. This last parameter is the first to be tested with a value of 4, 8, 12 and 16 over the three clustering methods.

Figure 2 shows the sensitivity of the clustering methods to the considered number of neighbors. First, the Average method gathers almost all the profiles in one cluster so it seems irrelevant for the classification. Second, the sensitivity to this parameter is very low for the Ward and Average methods but larger for the Spectral method. The classifications using the 16 nearest neighbors are distributed more equally between the clusters. Thus, for the following investigation, the Ward method is tested with 16 neighbors. For both areas, the classification is made for 2 to 10 clusters.

Figures 3 and 4 illustrate the different classifications made using different numbers of clusters. For the western equatorial Atlantic, the variability of the density profiles is mainly controlled by the depth of a single pycnocline. For 3 to 7 clusters, a

[1]https://scikit-learn.org/stable/modules/clustering.html

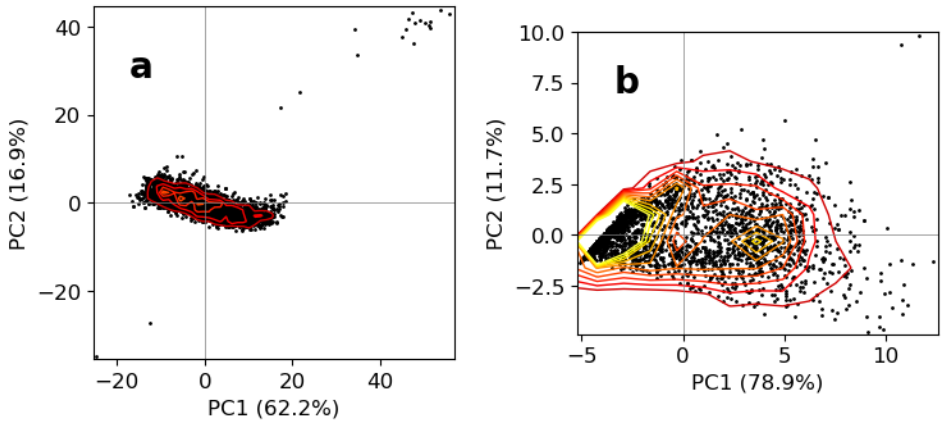

Figure S1. PCA manifold for (a) the western equatorial Atlantic and (b) the Bay of Biscay. The colored contours correspond to the density of points from 10 to 100 by tens.

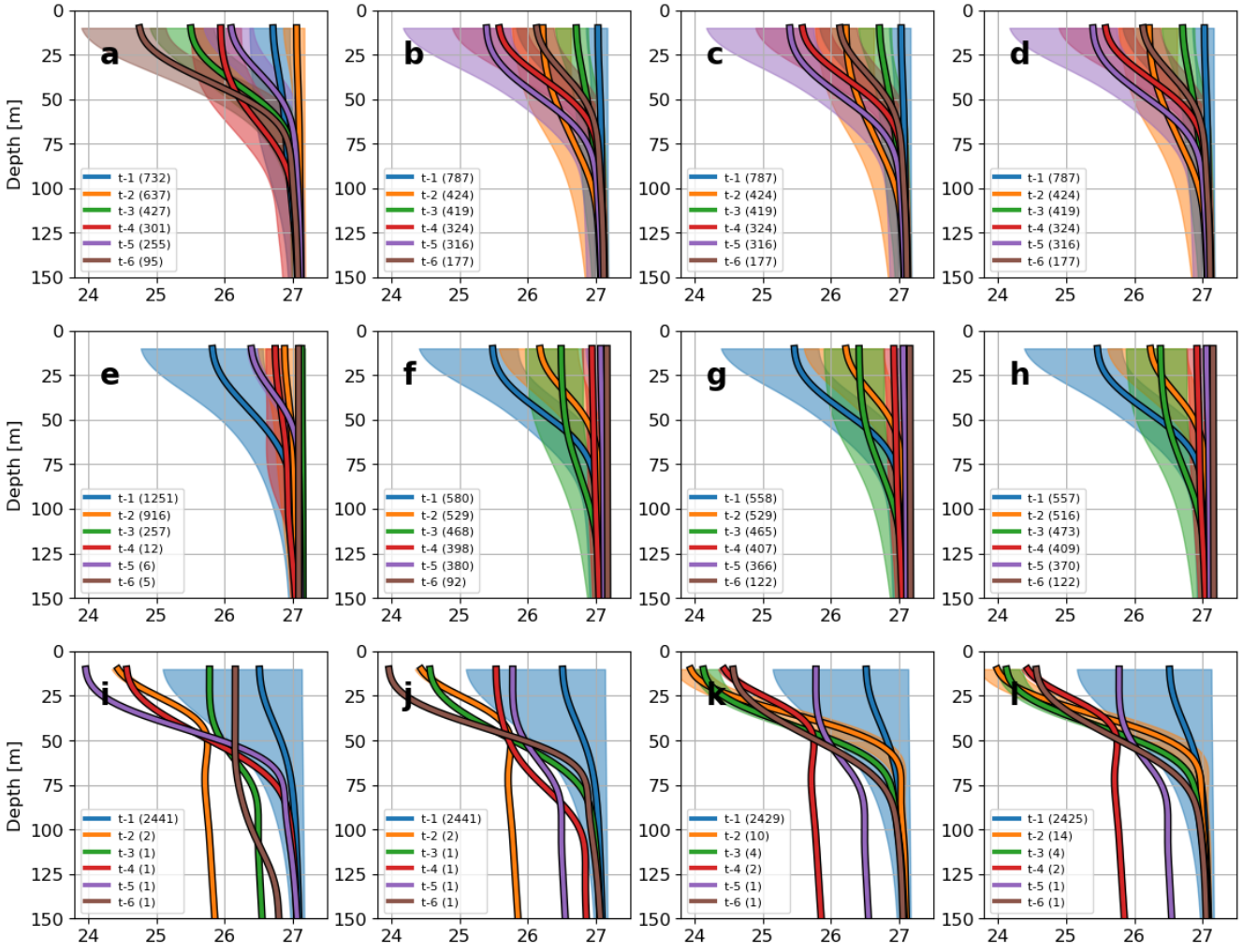

Figure S2. Clustering of the density profiles of the Bay of Biscay for (a-d) Ward, (e-h) Spectral and (i-l) Average methods. For each method, the calculation is made considering (a,e,i) 4 neighbors, (b,f,j) 8 neighbors, (c,g,k) 12 neighbors and (d,h,l) 16 neighbors. The number of profiles inside each cluster is given in the legend between brackets.

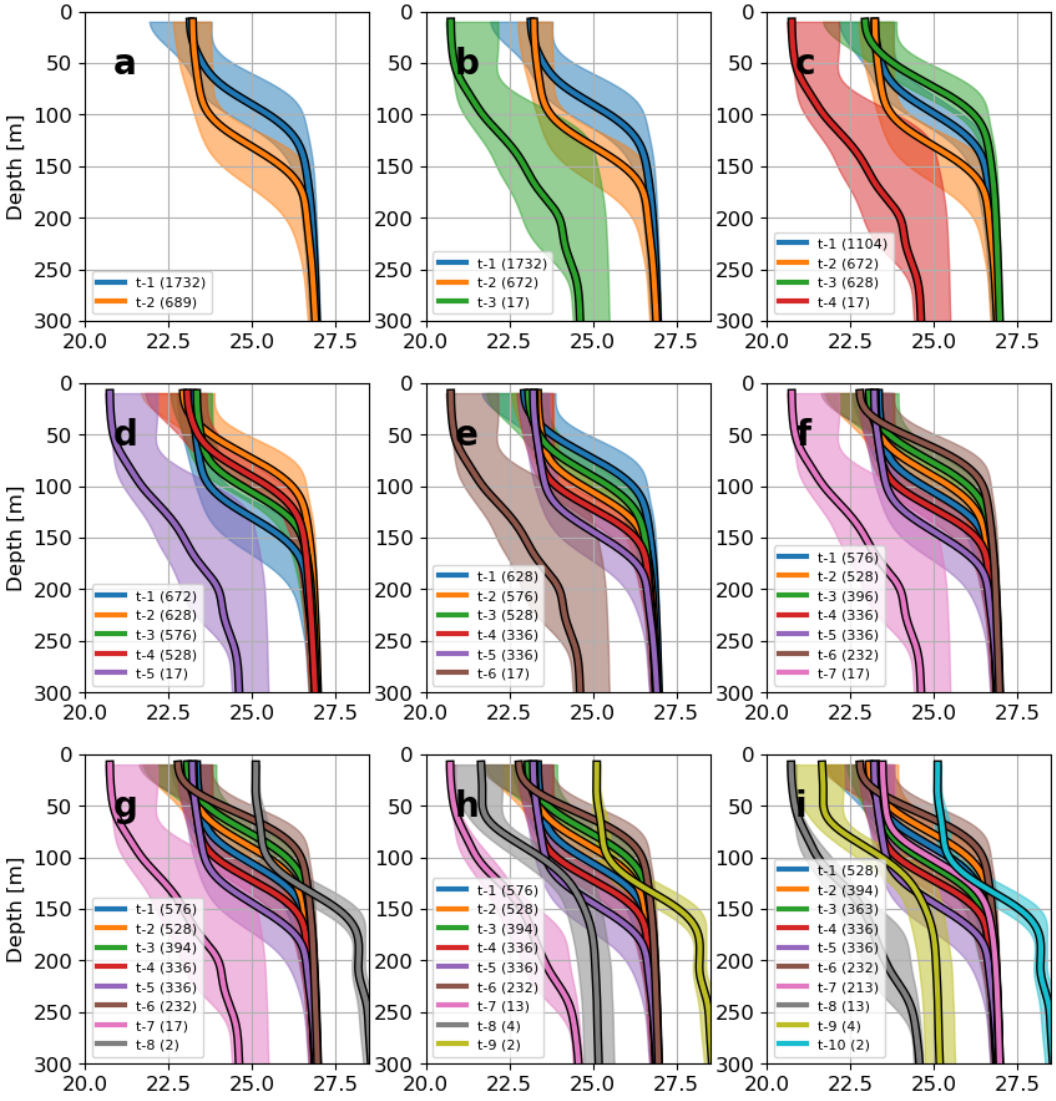

Figure S3. Ward clustering of the density profiles of the western equatorial Atlantic for a different number of clusters, from 2 to 10. The number of profiles inside each cluster is given in the legend between brackets.

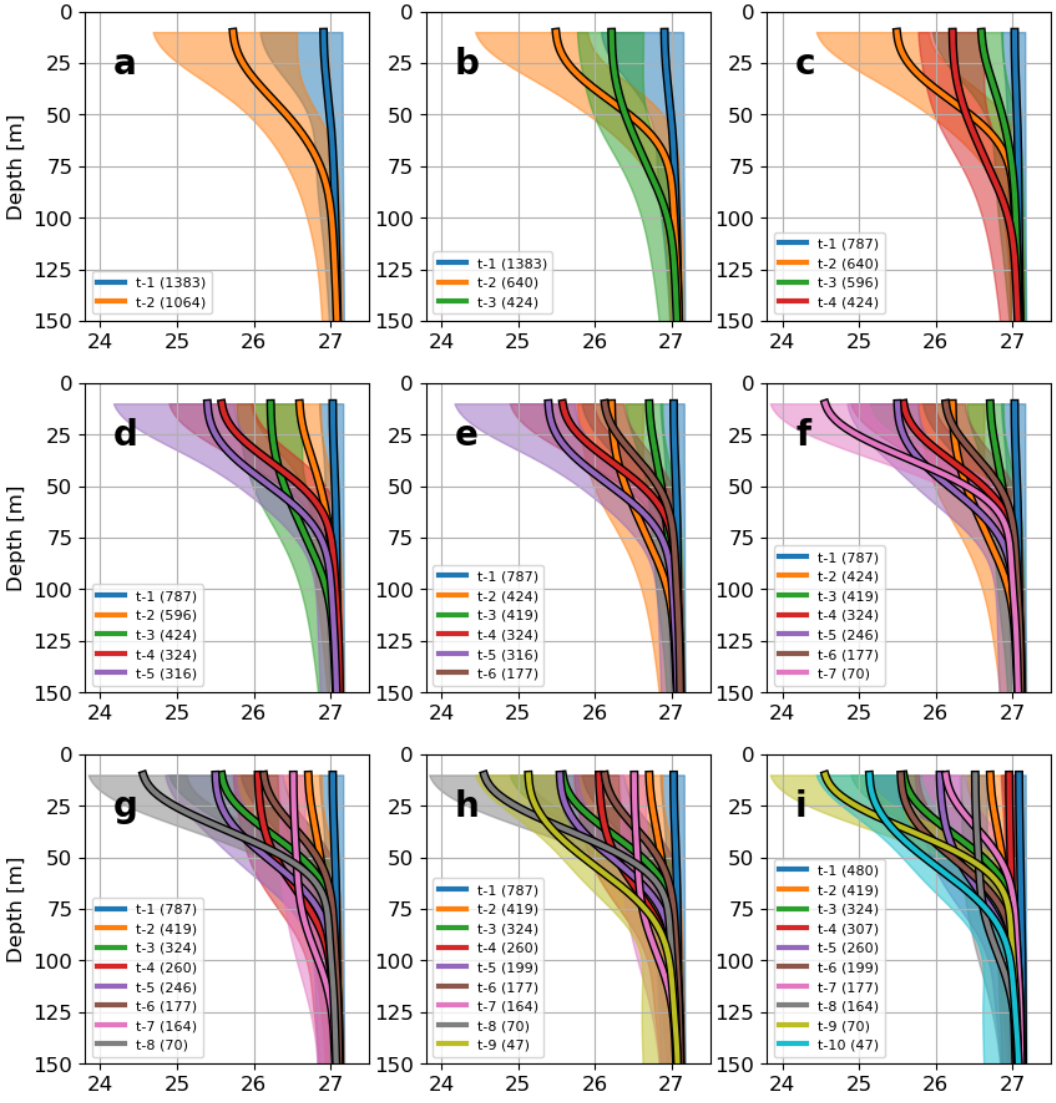

Figure S4. Ward clustering of the density profiles of the Bay of Biscay for a different number of clusters, from 2 to 10. The number of profiles inside each cluster is given in the legend between brackets.

unique cluster gathers suspicious profiles that seem to have an offset problem. For 8 to 9 clusters, the new clusters describe the different sets of suspicious profiles. For 10 clusters, the new clusters describe the main variability of the density profiles. The clusters of clean data are very similar and the increase of the number of cluster only increase the number of depth described for the single pycnocline. Thus, a few clusters can describe the full variability of the density profiles. For the Bay of Biscay, every cluster describes a different type of density profile that has a different surface density and a different N profile. But for 7 to 10 clusters, some clusters only describe a few profiles: less than 100 profiles over around 15 years of data, which questions the significance of these clusters.

References

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Sokal, R. R.: A statistical method for evaluating systematic relationships, Univ. Kansas, Sci. Bull., 38, 1409–1438, 1958.

Von Luxburg, U.: A Tutorial on Spectral Clustering, https://doi.org/10.1.1.165.9323, 2007.

Ward, J. J. H.: Hierarchical grouping to maximize payoff, Tech. rep., PERSONNEL RESEARCH LAB LACKLAND AFB TX, 1961.

Ward, J. J. H. and Hook, M. E.: Application of an hierarchical grouping procedure to a problem of grouping profiles, Educational and Psychological Measurement, 23, 69–81, 1963.

Yu, S. X. and Shi, J.: Multiclass Spectral Clustering, International Conference on Computer Vision, Nice, France, 11-17 Oct 2003, https://www1.icsi.berkeley.edu/~stellayu/publication/doc/2003kwayICCV.pdf, 2003.