Dynamic clustering for the identification of homogeneous acoustic landscapes

Postdoctoral position proposed by V. Audigier, F. Bouhadjera and N. Niang

CNAM, CEDRIC laboratory, MSDMA research team

2 rue Conté, 75003 Paris

Keywords: Dynamic clustering, functional data, heterogeneous data, high dimensional data, multiblock analysis, oceanography.

1 Introduction

Operational ocean forecasting aims at modelling the spatio-temporal evolution of the main physical and chemical properties of the ocean, including temperature, density, currents, and wave height. For underwater acoustics, the quantity of interest is the sound speed in seawater, which is derived from forecasts of temperature, density, and salinity through an equation of state. The elementary observations considered in this project are geo-referenced vertical sound-speed profiles measured at standard depth levels. These vertical profiles define acoustic waveguides that govern sound propagation in the ocean.

The volume of information generated by operational ocean models is often too large for applications constrained by limited communication bandwidth. This project therefore focuses on constructing reduced representations of these data. The problem can be formulated as a clustering problem, where the statistical units correspond to geographical locations (latitude $\times$ longitude) described by a data matrix denoted by X_{pred}. Each observation is characterized by a vector of scalar values (which may naturally be represented as a functional trajectory, see Figure 1), together with its latitude, longitude and time t. Rather than transmitting the complete forecast profile by profile, the objective is to transmit only a limited number of representative profiles associated with homogeneous clusters, together with their spatial and temporal domains of validity. Such problems

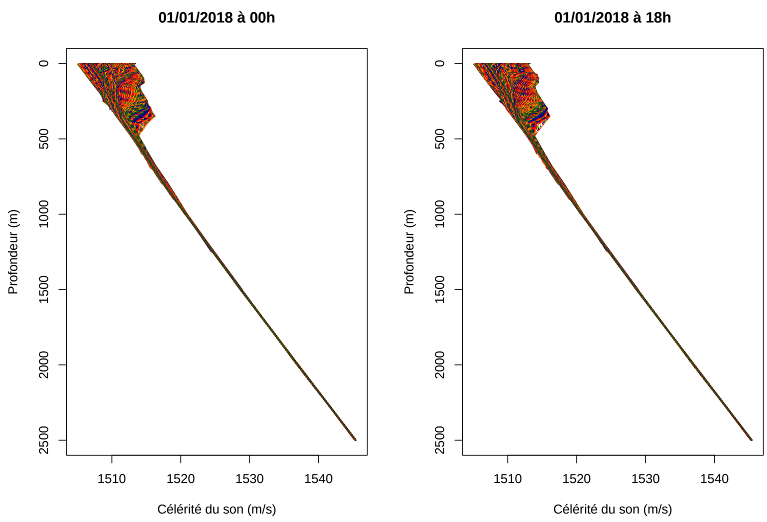

Figure 1: Examples of sound-speed profiles as a function of depth on January 1st, 2018 at 00:00 and 18:00.

can be addressed using unsupervised learning techniques, including Deep Neural Networks (DNN), k-means clustering, Self-Organizing Maps (SOM), and related approaches. However, these methods rely solely on the geometric similarity of the sound-speed profiles and therefore do not guarantee that similar profiles will exhibit similar acoustic propagation properties. Additional discriminative acoustic quantities can be associated with each sound-speed profile, such as channel cut-off frequency, acoustic wave resurfacing distance, or transmission loss as a function of horizontal range. It is therefore natural to suppose that jointly clustering the sound-speed profiles contained in X_{pred} together with their associated acoustic descriptors, denoted by Y_{pred}, would produce more meaningful and operationally relevant classifications.

The objective of this project is consequently to develop reduced representations based on clustering methods that incorporate not only oceanographic variables but also their associated acoustic descriptors. The resulting clusters will be interpreted as homogeneous acoustic landscapes.

2 Objectives

Homogeneous acoustic landscapes are constructed from a finite set of predicted ocean states, which provide only a partial description of the marine environment. An important question is therefore their spatial and temporal stability under the influence of ocean dynamics. The aim of this project is to explicitly account for the temporal dimension in the monitoring and updating of homogeneous acoustic landscapes, assumed to have been obtained through previous work. The project is organized into two complementary research tasks.

2.1 Stability analysis of the structure and composition of acoustic landscapes

The first task consists in investigating the temporal evolution of acoustic landscapes. This study may rely on multiblock data analysis methods (e.g., [1]), functional data analysis techniques for georeferenced data (e.g., [2]), functional time series analysis (e.g., [3]), or methods for categorical functional data analysis (e.g., [4, 5]).

The objective will first be to characterize the evolution of acoustic landscapes over time. Building on these results, a forecasting methodology will be developed to predict future acoustic landscapes from historical observations. Finally, idealized scenarios based on simulated datasets will be investigated to identify the mechanisms leading to major structural changes in acoustic landscapes, such as the emergence, disappearance, or merging of acoustic regions as ocean conditions evolve.

2.2 Adaptive updating of homogeneous acoustic landscapes

Representing the ocean environment by homogeneous acoustic landscapes inevitably results in a loss of information compared with the original ocean forecast. The purpose of adaptive (or onboard) updating is to compensate for this loss by incorporating additional information available to the end user. Such information may consist of in situ observations (X_{obs}, Y_{obs}) and prior knowledge or historical databases (X_{stat}, Y_{stat}).

The objective of this task is to develop onboard exploitation methods for acoustic landscapes at three complementary levels:

- **Reconstruction:** Reconstruction aims at recovering an environmental model (X_{pred}, Y_{pred}) from the compressed acoustic landscape representation. Particular attention will be paid to stochastic reconstruction methods exploiting class membership probabilities, within-class variability, and possibly topological relationships between neighboring clusters.

- **Onboard reliability assessment:** This task aims at determining the type, quantity, and acquisition strategies of in situ observations required to provide statistically reliable validation of the reconstructed environmental model. Particular emphasis will be placed on adapting the reconstruction process to representation uncertainty, with the objective of reducing uncertainty where it is most significant.

- **Updating:** Updating consists in constructing an improved environmental forecast (X_{mod}, Y_{mod}) by combining forecast information (X_{pred}, Y_{pred}), in situ observations (X_{obs}, Y_{obs}), and prior knowledge (X_{stat}, Y_{stat}). This work will focus on the development and application of data fusion and data assimilation techniques specifically adapted to the homogeneous acoustic landscape framework.

3 Provisional schedule

The postdoctoral position is expected to start within two months following the selection of the successful candidate. The project is organized into the following steps:

- Data handling and investigation of the temporal stability of acoustic landscape structure and composition (6 months).

- Development of methods for the adaptive updating of homogeneous acoustic landscapes (9 months).

Each research task is expected to lead to a scientific publication including an application to oceanographic data. In addition, the postdoctoral researcher will be expected to ensure the reproducibility of all developed methods and to prepare a final report summarizing the work carried out during the project.

4 Research environment

The selected candidate will join the CEDRIC laboratory at CNAM and will benefit from the laboratory's computational resources as well as the scientific expertise of the research team MSDMA.

5 Candidate profile

Applicants should hold a PhD in mathematics, statistics, data science, or a closely related field. Particular consideration will be given to candidates with expertise in functional data analysis, multiblock data analysis, or unsupervised learning. A strong background in statistical learning, data analysis, and scientific programming is expected. Candidates should also demonstrate excellent scientific writing skills in both French and English. The position requires regular on-site presence at CNAM in Paris.

6 Application procedure

Applications should include:

- a detailed curriculum vitae highlighting the candidate's suitability for the project;
- a cover letter explaining the candidate's motivation and research interests;
- the PhD thesis manuscript together with a list of publications and, where possible, copies of the main publications;
- letters of recommendation.

Applications should be sent by email to all three of the following addresses: vincent.audigier@cnam.fr, feriel.bouhadjera@lecnam.net, ndeye.niang2@lecnam.net.

Applicants will be informed of the outcome of their application in early September 2026.

References

[1] Lingjing Jiang, Chris Elrod, Jane J Kim, Austin D Swafford, Rob Knight, and Wesley K Thompson. Multi-block sparse functional principal components analysis for longitudinal microbiome multi-omics data. *arXiv preprint*, 2021.

[2] Alvaro Alexander Burbano-Moreno and Vinícius Diniz Mayrink. Spatial functional data analysis: Irregular spacing and bernstein polynomials. *Spatial Statistics*, 60:100832, 2024.

[3] Israel Martínez-Hernández and Marc G Genton. Functional time series analysis and visualization based on records. *Journal of Computational and Graphical Statistics*, 34(1):72–83, 2025.

[4] Cristian Preda, Quentin Grimonprez, and Vincent Vandewalle. Categorical functional data analysis. the cfda r package. *Mathematics*, 9(23):3074, 2021.

[5] Hervé Cardot and Caroline Peltier. Statistical modeling of categorical trajectories with multivariate functional data approaches. In *International Workshop on Functional and Operatorial Statistics*, pages 101–108. Springer, 2025.

[6] Charles Bouveyron, Julien Jacques, Amandine Schmutz, Fanny Simoes, and Silvia Bottini. Co-clustering of multivariate functional data for the analysis of air pollution in the south of france. *The Annals of Applied Statistics*, 16(3):1400–1422, 2022.

[7] Boubacar Diallo, Ndèye Niang, Feriel Bouhadjera, and Vincent Audigier. Classification de données fonctionnelles et vectorielles. In *57èmes Journées de Statistique (JdS)*, Clermond-Ferrand, France, June 2026. Société Française de Statistique.

[8] Yosra Ben Slimen, Julien Jacques, and Sylvain Allio. Co-clustering for binary and functional data. *Communications in Statistics - Simulation and Computation*, 2020.

[9] Gholamreza Soleimani and Masoud Abessi. Dlcss: A new similarity measure for time series data mining. *Engineering Applications of Artificial Intelligence*, 92:103664, 2020.

[10] Julien Jacques and Cristian Preda. Functional data clustering: a survey. *Advances in Data Analysis and Classification*, 8:231–255, 2014.

[11] Jeong Hoon Jang. Principal component analysis of hybrid functional and vector data. *Statistics in medicine*, 40(24):5152–5173, 2021.

[12] Mimi Zhang and Andrew Parnell. Review of clustering methods for functional data. *ACM Transactions on Knowledge Discovery from Data*, 17(7):1–34, 2023.