

Classification dynamique pour l'identification de paysages acoustiques homogènes

Sujet de postdoctorat proposé par V. Audigier, F. Bouhadjera et N. Niang

CNAM, Laboratoire **CEDRIC**, équipe **MSDMA**

2 rue Conté, 75003 Paris

Mots-clés : Classification non-supervisée dynamique, données fonctionnelles, données hétérogènes, données massives, analyse multiblocs, océanographie.

1 Introduction

La prévision océanographique opérationnelle consiste à modéliser le champ spatio-temporel des différents paramètres physiques et chimique de l'océan : température, densité, courants, hauteurs de vagues. Pour le besoin de l'acoustique sous-marine, la grandeur d'intérêt est la vitesse du son dans l'eau, qui se calcule à partir des prévisions de température, de densité et de salinité par une équation d'état. Les grandeurs élémentaires auxquelles s'applique cette recherche sont des profils verticaux de célérité géo-référencés à des profondeurs standards. Ces profils verticaux définissent des guides d'ondes qui régissent la propagation acoustique dans les océans.

Les volumes d'information générés par les modèles océaniques sont trop importants pour être utilisés dans certaines applications contraintes par la transmission de petits volumes de données. On s'intéresse dans cette recherche à obtenir des représentations réduites. On peut poser le problème de représentation géostatistique réduite comme un problème de classification d'une population d'individus (couples latitude $\times$ longitude) à partir d'un tableau de données noté X_{pred} . Chaque individu est défini par un vecteur de scalaires (qui peut être représenté sous la forme d'une trajectoire, voir Figure 1), par une latitude, par une longitude et par un temps t . Plutôt que de transmettre l'ensemble de la prévision profil par profil, on envisage de ne transmettre que certains profils caractéristiques d'un groupe homogène ainsi que leur domaine d'extension géographique et temporelle.

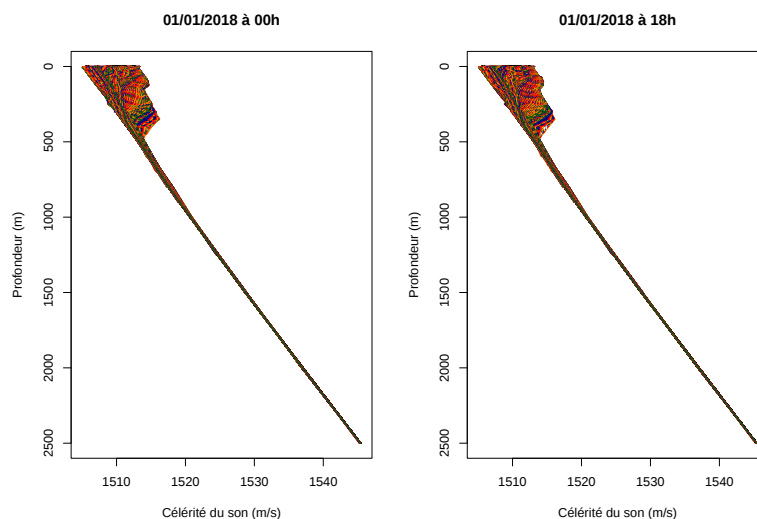


FIGURE 1 – Exemple de profils de célérité en fonction de la profondeur au 01/01/18 à 00h et 18h.

Ce type de problème peut être résolu par des algorithmes de classification non-supervisée (réseaux de neurones profonds, K-means, cartes topologiques, etc). Cependant, ces approches reposent seulement sur la forme des profils, ce qui ne garantit pas que des profils proches aient un comportement proche du point de vue de la propagation acoustique. Il est possible d'associer aux profils verticaux de célérité des grandeurs acoustiques discriminantes comme par exemple la fréquence de coupure de

chenaux, la distance de résurgence des ondes acoustiques, des pertes de propagation en fonction de la distance horizontale, etc. Il est intuitif de penser qu'une classification conjointe des profils verticaux à partir de X_{pred} et des variables acoustiques associées, que l'on désigne par Y_{pred} , pourrait améliorer la pertinence de la classification.

Ainsi, le but de cette recherche est de définir des méthodes de représentations réduites, reposant sur des techniques de classification, qui ne soient pas seulement fondées sur des analyses de variables océanographiques mais qui inclut aussi des variables acoustiques associées. Les groupes établis par ces méthodes pourront être interprétés comme des paysages acoustiques homogènes.

2 Objectifs

Les paysages acoustiques homogènes sont constitués à partir d'une connaissance partielle de l'environnement marin, au travers d'un nombre fini de situations prédites. Se pose la question de leur stabilité spatiale et temporelle sous l'effet de la dynamique océanique. Ce travail a pour but de prendre la dimension temporelle dans le suivi et l'actualisation des paysages acoustiques homogènes, que l'on suppose définis par un travail préalable. Le travail sera divisé en deux sous-tâches.

2.1 Étude de stabilité de la structure et de la composition des paysages sonores

Un premier travail consistera à explorer l'évolution des paysages acoustiques. On pourra pour cela recourir à des approches multiblocs (e.g. [1]), à des méthodes d'analyse fonctionnelle pour données géolocalisées (e.g. [2]), à des méthodes d'analyse des séries temporelles fonctionnelles (e.g. [3]) ou encore à des méthodes d'analyse de données fonctionnelles qualitatives (e.g. [4, 5]). Il s'agira ensuite de proposer une méthode de prévision du paysage acoustique à partir d'un historique. Enfin, on étudiera dans des cas idéalisés (i.e sur des échantillons réduits de données simulées) les conditions de changements radicaux de paysages sonores (disparition, apparition, regroupement) sous l'effet de l'évolution des situations prédites.

2.2 Perception et actualisation nomade des paysages acoustiques homogènes

La réduction en paysages sonores homogènes entraîne de fait une perte d'information par rapport à la prévision initiale. L'intérêt de l'actualisation nomade est de compenser cette perte d'information grâce à de l'information disponible chez l'exploitant, information qui peut prendre deux formes : des observations in situ (X_{obs}, Y_{obs}) et des bases de connaissances (X_{stat}, Y_{stat}). Le but de cette tâche est de définir les méthodes d'exploitation embarquées des paysages acoustiques à plusieurs niveaux :

- La reconstruction : la reconstruction consiste à obtenir un modèle d'environnement (X_{pred}, Y_{pred}) issu de la décompression de la représentation en paysages sonores. On s'intéressera notamment à obtenir une représentation stochastique de l'environnement en se basant sur l'exploitation des probabilités d'appartenance à une classe et la dispersion des classes, voire des relations topologiques entre classes voisines.
- La fiabilisation embarquée : la fiabilisation embarquée consiste à définir le type, le nombre ou encore les scénarios d'acquisition de données in situ qui offrent une garantie de pertinence statistique pour valider le modèle reconstruit. Cela passe en particulier par la recherche d'adaptation aux incertitudes de représentations ; le but des mesures et informations embarquées étant de réduire les incertitudes dans les situations où celles-ci sont importantes.
- L'actualisation : l'actualisation consiste à construire une prévision d'environnement (X_{mod}, Y_{mod}) en fonction des informations ($X_{pred}, Y_{pred}, X_{obs}, Y_{obs}, X_{stat}, Y_{stat}$). Ce travail portera sur la définition et l'application des outils de fusion et d'assimilation de données les plus adaptés au contexte de la représentation par paysages acoustiques homogènes.

3 Calendrier prévisionnel

Le postdoctorat pourra débuter dans les deux mois succédant la sélection du candidat. Le travail se déroulera selon les étapes suivantes :

- Prise en main des données et étude de stabilité de la structure et de la composition des paysages sonores (6 mois).
- Perception et actualisation nomade des paysages acoustiques homogènes (9 mois).

Chacune des propositions fera l'objet d'une publication intégrant une application à des données océanographiques.

Par ailleurs, le postdoctorant devra fournir un travail reproductible et rédiger un rapport d'activité résumant les travaux effectués.

4 Moyens consacrés

Les moyens tant matériels (informatiques) qu'humains (compétences propres) du laboratoire d'accueil (CEDRIC/CNAM) seront mis à disposition du doctorant.

5 Profil du candidat

Candidat disposant d'un doctorat dans le domaine des mathématiques, de la statistique, ou de la science des données. Une attention particulière sera portée aux candidat ayant une expérience dans l'analyse des données fonctionnelles ou multiblocs. Un bon niveau en analyse des données, en programmation, ainsi que des capacités à rédiger en Français et en Anglais sont attendues. Nécessité d'être présent sur le site.

6 Modalités de candidature

Les dossiers de candidatures devront être composés

- d'un cv détaillé, présentant l'adéquation du candidat par rapport au sujet
- d'une lettre de motivation mettant en évidence les raisons de la candidature
- du rapport de thèse et des différentes publications du candidat
- de lettres de recommandation

Ces éléments devront être transmis par mail aux trois adresses suivantes : vincent.audigier@cnam.fr; feriel.bouhadjera@lecnam.net; ndeye.niang2@lecnam.net.

Les candidats seront informés de la suite donnée à leur candidature au début du mois de septembre 2026.

Références

- [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, Ferial 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.