ÉCOLE DOCTORALE

SCIENCES DE LA TERRE ET DE L'ENVIRONNEMENT ET PHYSIQUE DE L'UNIVERS, PARIS

ed560.stepup@u-paris.fr

Title : Uncovering seismic signals in submarine DAS data with interpretable machine learning

Directeur:
METAXIAN Jean-Philippe, (IRD IPGP), metaxian@ipgp.fr

Co-encadrant :
VIENS Loïc, (LALN, NM, USA), lviens@lanl.gov

Equipe d'accueil :
IPGP - Équipe de sismologie – UMR7154
LANL

Financement : **Financements LANL + ASN**

Développement du sujet : (Maximum 2 pages)

Scientific context

Submarine environments, covering more than 70% of Earth's surface, represent one of the final frontiers for Earth observation due to their inherent remoteness and technical challenges. The lack of instrumentation severely limits our ability to monitor natural hazards, such as earthquakes, volcanic eruptions, and tsunamis, as well as anthropogenic activities in these regions. Distributed Acoustic Sensing (DAS) represents a transformative innovation by turning existing underwater telecommunication fiber-optic cables into dense and cost-effective seismic arrays. DAS offers unprecedented spatial resolutions (meter-scale) and continuous temporal monitoring capabilities across hundreds of kilometers. However, DAS also introduces new challenges, including extremely high data volumes, non-stationary signal characteristics, and a lack of prior labeling, particularly in submarine environments. The scientific community increasingly recognizes DAS's potential across a broad range of geophysical applications, from seismic imaging and volcano monitoring to marine ecology and environmental science. Despite its promise, DAS data analysis remains in its infancy, necessitating the development of advanced methods capable of handling the scale, complexity, and diversity of the recordings. Conventional processing methods and supervised machine learning models are often insufficient, highlighting the critical need for novel, interpretable, and scalable approaches.

Objectives

The objective of this PhD project is to design, implement, and apply interpretable machine learning methods, specifically 2D scattering networks, to unlock information embedded in submarine DAS recordings. Scattering networks, based on cascaded wavelet transforms, produce multi-scale, translation-invariant representations of structured data, making them ideal for processing spatiotemporal DAS datasets where coherent signal patterns are often hidden within noisy backgrounds. Unlike black-box deep learning models, scattering networks provide mathematical guarantees of stability and interpretability, critical for scientific inference.

Key goals include:

- Developing tailored 2D scattering network architectures adapted to the unique characteristics of submarine DAS data.
- Combining scattering features with unsupervised learning techniques to automatically detect, classify, and characterize natural and anthropogenic signals.
- Achieving robust signal detection across variable coupling conditions and noise environments.
- Producing interpretable representations that can provide new insights into the physics of submarine seismic and environmental processes.

The project will leverage two unique datasets:
- A continuous offshore DAS dataset from Cordova, Alaska (LANL), one of the longest-duration submarine DAS recordings globally, with extensive seismic, icequake, and anthropogenic signals.
- A submarine volcanic DAS dataset from Santorini, Greece (IPGP), collected during a period of volcanic unrest, offering a unique window into seismo-volcanic activity.

Methodology
The research plan is organized as follows:
- Design and implement 2D scattering transforms specifically adapted to spatiotemporal DAS "images." Exploration of different wavelet configurations and scales to optimize feature extraction.
- Apply of dimensionality reduction and clustering to identify and classify signal types.
- Test and benchmark on synthetic DAS datasets with known characteristics to rigorously evaluate performance and interpretability.
- Appy of methods to Cordova and Santorini datasets, with careful domain adaptation strategies to bridge the gap between synthetic and real-world data.
- Integrate external data sources, including earthquake catalogs, ship tracking data, tidal records, and environmental data, to validate and interpret clustering results.

Workplan
- Task I: Development and rigorous testing of 2D scattering network architectures on synthetic datasets.
- Task II: Large-scale application to Cordova DAS dataset; creation of an annotated catalog of natural and anthropogenic signals; detailed analysis of signal characteristics.
- Task III: Application to Santorini DAS dataset; characterization of volcanic tremor, earthquake swarms, and background seismic noise evolution during the ongoing volcanic crisis.

The PhD candidate will alternate between IPGP and LANL, spending approximately five months per year at LANL. Weekly or biweekly meetings between teams will ensure seamless collaboration.

Expected outcomes
- Development of advanced and interpretable machine learning algorithms tailored to DAS data.
- New methodologies for detecting and interpreting seismic and environmental processes in submarine settings.
- High-impact open-access publications and signal catalogs contributing to the broader geophysical community.
- Strong interdisciplinary training at the interface of geophysics, applied mathematics, and machine learning.

Candidate profile
We seek a highly motivated candidate with a strong background in signal processing, machine learning, geophysics, applied mathematics, or a related discipline. Proficiency in Python programming and experience with scientific computing libraries are expected. Prior exposure to wavelet analysis, unsupervised learning, seismic data processing, or high-performance computing is a plus. The ideal candidate should demonstrate analytical rigor, scientific curiosity, and enthusiasm for interdisciplinary research in international settings.

Collaborators

- Léonard Seydoux (Pr, IPGP), seydoux@ipgp.fr