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

ed560.stepup@u-paris.fr

**Subject title:** **Rebuilding past seismic catalog with machine learning: how far can we go?**

Advisor: **RETAILLEAU Lise, physicienne adjointe, retailleau@ipgp.fr**

Co-advisor: **SEYDOUX Léonard, Pr, seydoux@ipgp.fr**

Research team: **IPGP- Equipe de sismologie – UMR7154**

Financement : **Contrat doctoral avec ou sans mission d'enseignement**

**Développement du sujet :** (Maximum 2 pages)

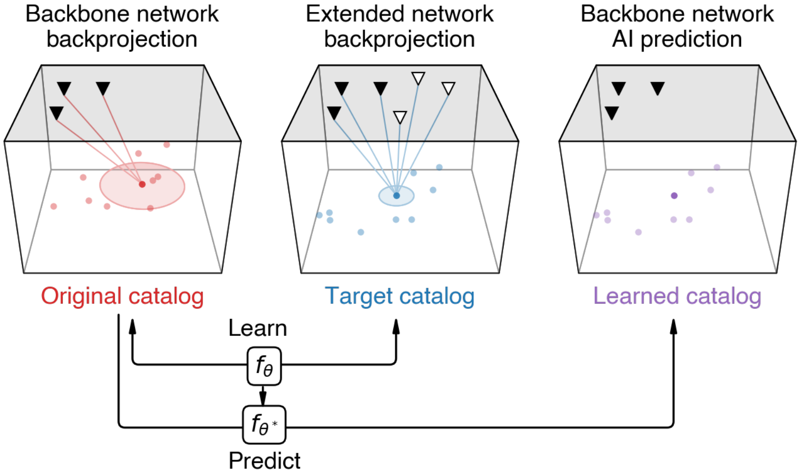
**Context and motivation**

Earthquakes express stress release in active geological objects and provide geophysicists with crucial information about the underlying physical processes, such as fluid migrations, stress perturbations, and incoming events like volcanic eruptions (Chouet and Matoza, 2013). Forming accurate seismic catalogs is therefore essential to monitor and understand these systems. The quality of seismic catalogs relies on the detection capacity and location accuracy. The detection capacity depends on the background seismic noise generated by either anthropogenic activity or oceanic activity. This capacity strongly improved thanks to the use of deep learning (e.g., Zhu & Beroza; 2019). Accurate localisation remains a challenge. It strongly depends on the geometry of the seismic networks, such as the number of stations, stability, and geographical location.

Since the 2000s, seismologists have densified and extended seismic networks (Arrowsmith et al., 2022) which has led to build high-quality catalogs. Historically, this trend has been primarily triggered by the occurrence of major events such as volcanic eruptions or large earthquakes (e.g., Romanowicz & Dziewonski, 2010). For example, seismic networks are usually deployed permanently or temporarily in the vicinity of these events, where only a few stations were available, to study in detail the subsequent seismic activity which provides strong insights into the main event (e.g., Rietbrock et al., 2012). As a consequence, the main activity, and potential precursors, are usually not recorded by the dense deployments, preventing a full understanding of the event's dynamics.

This is the case of the large eruption of the Fani Maoré volcano in Mayotte which was first measured by one local seismic station. The deployment of ocean-bottom seismometers (OBS) and land seismic sensors started months after the beginning of the pre-eruptive seismic swarm (Saurel et al., 2022). It significantly improved the understanding of the seismic activity linked to the volcanic system but the first and most active activity was only poorly recorded.

In this project, we propose to leverage the knowledge gained from the dense deployments to infer the location of events from a single or a few long-term seismic stations. To manage this, we will use machine learning strategies. We will target the eruption of Fani Maoré volcano, as it is an ideal test case to develop the method. The proposed solutions consider several levels of complexity that the successful candidate will evaluate with increasing complexity until a result reaches sufficient accuracy.




**Proposed implementation**

The goal of this project is to correct the systematically biased earthquake locations obtained with a permanent and sparse seismic network (backbone network), from the locations inferred with a denser deployment with adequate geometry (extended network), as illustrated in the figure. We plan to solve the problem with supervised machine learning, where the input data and output data have a variable complexity depending on the target accuracy, and the backbone network quality.

The student will build on the proof-of-concept work of Mohammadi et al. (in prep.), which used a random forest regression technique out of the location likelihood. While using the likelihood to infer the location led to promising results, the strong inhomogeneity of earthquake location yields to biased and potential overfitting. We plan to tackle these issues with additional features (e.g., arrival times, velocity models) to further improve the predicted location accuracy and robustness.

In a second stage, we will address the problem of single-station location prediction out of the seismograms directly deep learning (Mousavi and Beroza, 2020). This will address remaining questions about the processes during the magma propagation before the Fani Maoré eruption, our test case, as only one local station recorded at that time.

We expect this project to build earthquake catalogs with unprecedented spatial resolution in time by revisiting limited past data, and enable accurate analyses of past and present seismic activity. We will investigate various aspects of these catalogs, and infer other event characteristics such as magnitude or statistical properties. We will also coordinate with collaborators to interpret our results in light of their connex expertises (volcanology, petrology, etc.). Ultimately, we will consider other cases to challenge the method and broaden the impact of the project.

The candidate will work full time in the Cuvier site of the Institut de physique du globe de Paris (IPGP) in the seismology team. He will closely interact with Lise Retailleau and Léonard Seydoux during the entire project. The IPGP provides an environment of excellence for PhD students, with many inter-student and inter-teams interactions. The PhD candidate should have a strong interest in geophysics, seismology, signal processing, programming (Python preferred). A good background in statistics and machine learning is appreciated.

**References**


- Chouet, B. A., & Matoza, R. S. (2013). A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. Journal of Volcanology and Geothermal Research, 252, 108-175.
- Zhu, W., & Beroza, G. C. (2019). PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261-273.
- Romanowicz, B., & Dziewonski, A. (Eds.). (2010). Seismology and structure of the Earth: Treatise on Geophysics.
- Arrowsmith, S. J., Trugman, D. T., MacCarthy, J., Bergen, K. J., Lumley, D., & Magnani, M. B. (2022). Big data seismology. *Reviews of Geophysics*, *60*(2), e2021RG000769.
- Rietbrock, A., Ryder, I., Hayes, G., Haberland, C., Comte, D., Roecker, S., & Lyon-Caen, H. (2012). Aftershock seismicity of the 2010 Maule Mw= 8.8, Chile, earthquake: Correlation between co-seismic slip models and aftershock distribution?. Geophysical Research Letters, 39(8).
- Saurel, J. M., Jacques, E., Aiken, C., Lemoine, A., Retailleau, L., Lavayssière, A., ... & Feuillet, N. (2021). Mayotte seismic crisis: building knowledge in near real-time by combining land and ocean-bottom seismometers, first results. Geophysical Journal International, 228(2), 1281-1293.
- Retailleau, L., Saurel, J. M., Zhu, W., Satriano, C., Beroza, G. C., Issartel, S., ... & OVSM Team. (2022). A wrapper to use a machine-learning-based algorithm for earthquake monitoring. *Seismological Research Letters*, *93*(3), 1673-1682.
- Mousavi, S. M., & Beroza, G. C. (2020). Bayesian-Deep-Learning Estimation of Earthquake Location From Single-Station Observations. IEEE Transactions on Geoscience and Remote Sensing, 58(11), 8211-8224.