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

ed560.stepup@u-paris.fr

Subject title: Detection and classification of nanoparticles in spICP-ToF-MS time series using Machine/Deep learning algorithms

Advisor: **CHAUSSIDON, Marc, DR, chaussidon@ipgp.fr**
Supervisors:
THARAUD, Mickaël, IR, tharaud@ipgp.fr
Host lab/ Team:
IPGP – Teams CAGE, ACE – UMR 7154

Financing: Doctoral contract with or without teaching assignment

Scientific context

Environmental nanogeochemistry is emerging as a critical field for understanding contaminant dynamics within the Earth's critical zone. Nanoparticles (NPs), whether naturally occurring or anthropogenic, play key roles in the transport, bioavailability, and toxicity of trace elements. However, their quantification and identification in complex environmental matrices remain major analytical and methodological challenges, due to their low concentrations, high diversity, and the massive volume of data required. There is at present no statistically significant dataset allowing to estimate the relative abundances of various NPs in function of their various anthropogenic and natural origins in major urban areas. This is critically lacking to study the potential impacts of NPs on our environment.
Single-particle inductively coupled plasma mass spectrometry with time-of-flight analyzers (spICP-ToF-MS) now allows multi-elemental characterization of individual NPs with unprecedented resolution. However, processing such a set of analytical data (which can reach tens of millions of mass spectra in a single sample) faces several bottlenecks: imperfect peak detection, poor handling of coincident or saturated signals, and limitations of traditional clustering approaches such as hierarchical agglomerative clustering (HAC).

Thesis Objectives

This PhD project aims to produce for the first time a comprehensive data set on the origin and distribution of NPs in different zones of a heavily populated urban area exposed to various anthropic contributions. This will be reached by developing an innovative approach based on **Sequence-to-Sequence (Seq2Seq) deep learning models** to optimize the **detection, classification, and interpretation of NPs** within spICP-ToF-MS time series. By combining state-of-the-art artificial intelligence with specific challenges in environmental nanogeochemistry, this project will contribute to building a robust, reproducible, and transferable methodology for diverse environments, not only urban but also for tracing volcanic NPs or even looking for primitive dust components in extraterrestrial materials.
Specific objectives include:

- Developing Seq2Seq models tailored to spICP-ToF-MS time series to achieve better nanoparticle detection performance compared to classical thresholding or statistical methods;
- Comparing different architectures (BiGRU, BERT, TransNILM, U-Net, autoencoders) for handling noisy and complex signals;

- Identifying and classifying NP families using advanced unsupervised clustering techniques (DBScan, Spectral Clustering) to trace their origin (natural vs. anthropogenic);

Methodology and expected contributions

The PhD will be structured into several phases:

1. **Exploratory Phase**: Review of Seq2Seq architectures for complex time series analysis and initial testing on synthetic datasets.
2. **Development of Detection Pipeline**: Implementation and training of models on a controlled synthetic dataset, gradually introducing complexity (coincidences, saturation).
3. **Unsupervised Classification**: Exploration of clustering strategies to distinguish types of NPs based on their multi-elemental fingerprints.
4. **Study of samples of aerosols and rains**: Application of the developed models to spICP-ToF-MS data sets produced by the analyses of samples collected in different areas and at different periods in the chosen urban area. Production of a statistically significant survey of the distribution of NPs in the studied area.

The developed algorithms will be integrated into open-source tools, ensuring reproducibility and dissemination to the broader scientific community.

Work environment

This PhD will be part of the **nanoNET** project, funded by CNRS and led by an interdisciplinary team combining the Institut de Physique du Globe de Paris (IPGP), Université Gustave Eiffel, and École Normale Supérieure (ENS Paris). The PhD candidate will benefit from joint supervision by **Dr. Marc Chaussidon** and **Dr. Mickaël Tharaud**. The proposed project will be carried out at the Institut de physique du globe de Paris, in close collaboration with Pr. Leonard Seydoux (team Seismology) and Dr. Pierre Boniol (ENS), specialist in deep learning for time series. IPGP hosts a world-leading geochemistry platform (PARI), equipped with an operational spICP-ToF-MS instrument. We possess an extensive dataset of spICP-ToF-MS measurements for both engineered and natural NPs as well as glass standard NPs. This dataset will serve as the foundation for algorithm development and testing, before applying the methodology to the natural samples that will be collected during the PhD. Additionally, through various collaborations in France (Dr. Pierre-Emmanuel Peyneau, Université Gustave Eiffel), we can beneficiate from run-off water from an urban catchment that is part of the Service National d'Observation Observil (**nanoSTREAM** project) as well as simulate spICP-ToF-MS dataset for training and validation of the developed methodology.

Scientific background of the candidates. Applicants should hold a strong background in data science, artificial intelligence, signal processing, or environmental geochemistry with a clear interest in interdisciplinary research. Experience with Python and deep learning libraries (TensorFlow, PyTorch) is highly desirable. Strong motivation for handling large scientific datasets is expected. He/She should also demonstrate its teamwork capabilities and have good communications skills in English (written and oral).

Application. Application should be sent as soon as possible by email to the 3 supervisors. It should include:

1. A CV
2. A motivation letter (1 page maximum)
3. We also request at least one recommendation letter from previous supervisors

Interviews for pre-selected candidates will be made shortly after application deadline.