Title of the internship: Searching for the Unknown – anomaly finding in Large astronomical datasets

Supervisor : Emmanuel Gangler
Laboratory : LPC
University : Université Clermont Auvergne (France)
Email : emmanuel.gangler@clermont.in2p3.fr

Summary :

For long, the dynamics of the deep Universe has been only studied through static data : this was though the analysis of still images and spectra that the representation of a dynamic Universe emerged. Only a limited quantity of objects were known to present an intrinsic variable nature. However, with the advent of upcoming astronomical telescopes and facilities, new observation strategies become available, opening for systematic studies of everything that changes in the Universe.

With the unprecedented data volume acquired by the ZTF survey and soon by the LSST – one of the major facilities of this decade – discoveries of brand new type of astrophysical phenomena are almost certain. However, the challenge is now to distinguish them from billions of artifacts or objects belonging to already known classes. In other words, we have to find not only needles in a hay stack, but we also have no clue of what a needle looks like. Fortunately, this is a domain where machine learning shines.

The main goal of this M2 internship is to adapt machine learning tools developed in the general context of anomaly finding for time data to the particularities of the ZTF transient data set. Namely, adapting matrix profile techniques to this data stream appears to be a promising lead. The expected workplan for the internship is thus following: get a global understanding of the data and the scientific goal; establish similarity criteria and adapt the algorithmics to the specific case at hand, with the help of expertise from LIMOS computer researchers; run tests on the dataset with the help of the FINK broker, and hopefully report unknown anomalies to the scientific community. Applications of the skillset acquired during this internship would easily open the door for a continuation in astrophysics and/or data science PhD.

The successful candidate should be proficient with numerical and statistical tools, as well as be fluent in programming and using scientific libraries. A solid background in astrophysics, cosmology, and/or familiarity with modern coding environment would be a plus. The candidate is also expected to report their work to an international audience as well as the University interdisciplinary working group on Artificial Intelligence.