**Lieu du projet :** UFC, Institut UTINAM UMR6213 CNRS-UFC, équipe PhAs

**Titre de la thèse :** MODELING THE 3D MILKY WAY USING MACHINE LEARNING WITH GAIA AND INFRARED SURVEYS

**Encadrants de la thèse :** Annie Robin (HDR), 03 81 66 69 41 / Julien Montillaud , 03 81 66 69 25

**Descriptif des travaux de la thèse, indication des laboratoires concernés, planning des recherches du doctorant, lieu… :**

Understanding the 3D structure of the Milky Way is a pivotal step to address key problems in galaxy-scale astrophysics like the interplay between the Galaxy dynamics and star formation. This complex problem involves a large variety of physical processes from the chemistry of interstellar gas and dust to the dynamics of stars and interstellar matter in the gravitational potential of the Galaxy. Many large data sets have recently been or will soon be released that will enable a breakthrough in this field. The Gaia mission has already released distance measurements to ~2 million stars and ~1 billion is eventually expected. This will lead to 3D maps of stars and interstellar extinction with unprecedented resolution. However, through the Galactic plane the reach of this data set will be hampered by severe dust extinction.

Infrared (IR) surveys like 2MASS in the near-IR, GLIMPSE in the mid-IR and Herschel in the far-IR provide complementary data sets where both stars and dust can be observed much further than with Gaia, nevertheless without direct distance estimates. In nearby regions, a complete tomography of star formation regions can be achieved by the combination of Gaia distance and extinction measurements with these IR maps. However, their limited angular resolution prevents from resolving individual star formation regions at distances of several kiloparsec, and therefore makes it difficult to build a reliable model of the Milky Way structure at such distances.

Machine learning techniques have been successfully employed in many fields from character recognition to particle physics. The basic idea of machine learning is to “learn” approximate relationships between input and output data without any explicit analytical prescription being used. In astrophysics, it has proven efficient in situations relatively similar as the present one: estimating scalar physical properties (e.g. the mass of a galaxy) after training a code with high-resolution density structures extracted from hydrodynamical simulations. The PHD project consists in using nearby interstellar clouds and star forming regions as templates of similar but unresolved structures elsewhere in the Galaxy. The student will develop a code based on a machine learning technique. Selected nearby structures will be modelled in full details: the stellar content will be extracted from Gaia data; the gas 3D structure and the spectral energy distribution of dust emission will be computed using an available Monte-Carlo radiative transfer code, and constrained using both Gaia and IR data. These 3D templates will be used to simulate further structures seen from various angles and at various distances in order to “teach” the machine learning code.

The modelled 3D structures of the nearby clouds of the training sample will already provide a highly valuable data set for star formation studies. Obtaining a Milky Way 3D map of star formation regions thanks to machine learning will constitute a major step in understanding the effects of the Galactic environment on star formation.