ASTRO - The Galactic Genome Project -- stellar characterisation with the Gaia satellite and machine learning

Supervisors: Prof Jay Farihi & Dr Jason Sanders

The PhD project work will consist of three complementary work programmes. There is considerable scope for the student to contribute their own ideas and investigate more deeply the different areas. However, they are united by the use of a common dataset (which requires initial investment and effort to get up to speed with). The projects also range from solid plans that are guaranteed to yield results (through pilot studies by summer students) to more speculative projects that will require a more investigative approach paving the way for future Gaia data releases.

1. A catalogue of Milky Way carbon stars -- carbon stars can be produced by single star and binary channels so provide valuable tests for our understanding of stellar evolution and binary populations. The student will build a classifier to identify "all" of the carbon stars in the XP spectra dataset. This will involve the investigation of appropriate features from the spectra and the applicability of different machine-learning algorithms (e.g. random forests, neural networks). Furthermore, regressors will be investigated to see whether carbon abundances can be extracted for each star. The astrometry from Gaia for the sample will be investigated to assess binarity (through astrometric residuals) and the orbits (possibly complemented through follow-up programmes to obtain radial velocities). This approach has shown early promise through a summer student project.

2. Chemically peculiar stars -- the student will investigate supervised classification based on higher resolution spectroscopic catalogues e.g. APOGEE, GALAH, WEAVE to find evidence for unusual chemical enhancement. For example, towards the Galactic bulge, there are significant numbers of nitrogen-rich stars which were possibly formed in clustered star formation in the early Milky Way disc. This approach will be complemented by unsupervised classification to identify unusual and rare objects e.g. Sanders & Matsunaga (2023) have already demonstrated the power of UMAP on the LPV dataset.

3. The impact of variability on XP spectra -- in the current Gaia data release (DR3), the XP spectra are averages over multiple epoch observations. In this way, any spectral variability is washed out but possibly still detectable in the uncertainties. It has already been demonstrated that XP spectra have the power to detect polluted white dwarfs -- those contaminated by accretion of material from planets. It is therefore an interesting question whether pollution variability can be detected in these spectra. The student will investigate the behaviour of the XP spectra in the presence of variability. This project will naturally lead towards the availability of epoch XP spectra in Gaia DR4. Through these early investigations, the student will position themselves to be able to rapidly exploit this unusual dataset.