NASSP Master's Project Proposal 2026

Level of the project: Master's
Project title: Resolving Source Confusion in the MIGHTEE XMM-LSS Field using XID+
Primary supervisor: Dr. Eliab Malefahlo (University of South Africa),
Email: emalefe@unisa.ac.za or eliabmalefahlo3@gmail.com

Background and Motivation:
The current generation of radio telescopes, such as the MeerKAT telescope, has reached the sensitivity required to probe the μJy radio sky in unprecedented detail. At these faint flux limits, the radio sky is dominated by star-forming galaxies (SFGs) and radio-quiet Active Galactic Nuclei (AGN). However, deep radio surveys can be limited by source confusion. This occurs when the surface density of faint sources is so high that multiple physically distinct galaxies blend together within the telescope's synthesized beam. In this case, it becomes difficult to identify and separate individual objects, often leading to flux boosting, blending, and incomplete source recovery.

A promising solution is prior-driven probabilistic deblending. This technique uses higher-resolution multi-wavelength observations (e.g., optical and near-infrared) to provide accurate positional coordinates for galaxies. These are then used as priors when modelling the radio map, allowing us to estimate how much radio flux comes from each galaxy, even when several overlap within the telescope beam. We have recently developed and validated a deblending framework using the Bayesian tool `XID+` for the MIGHTEE-COSMOS field (Malefahlo et al. 2026).

Project Aim:
The aim of this project is to adapt and apply this successful probabilistic deblending framework to the much larger XMM-LSS field observed as part of the MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) survey. A key objective is to investigate a new strategy for constructing positional priors. The process begins by identifying and removing stars from the dataset, as they are unlikely to produce detectable radio flux. Following this, the student will use a stellar mass-limited selection on the remaining galaxies to build the target list. Stellar mass correlates strongly with star formation rate through the galaxy main sequence, making massive galaxies statistically more likely to host detectable radio emission or active galactic nuclei. This method provides a highly efficient way to identify radio-emitting sources. Ultimately, the student will produce a high-fidelity, deblended radio catalogue for XMM-LSS, enabling highly accurate measurements of faint radio source counts well beyond the traditional confusion limit.

Data:
The student will utilize state-of-the-art multi-wavelength datasets:

- **Radio Data:** Deep 1.3 GHz radio continuum maps from the MeerKAT MIGHTEE survey over the XMM-LSS field.
- **Multi-wavelength Data:** Deep optical and near-infrared catalogues covering the XMM-LSS field (e.g., from the HSC-SSP, VISTA-VIDEO, and SERVS surveys). These catalogues provide reliable photometric redshifts and stellar mass estimates derived via Spectral Energy Distribution (SED) fitting.

Methodology:
Building on the methodology established in Malefahlo et al. (2026), the student will:

1. **Star-Galaxy Separation:** Utilize unsupervised machine learning techniques (UMAP dimensionality reduction coupled with HDBSCAN clustering) on the multi-wavelength photometry to accurately separate and remove stars from the dataset.
2. **Prior Construction via Stellar Mass:** Apply a stellar mass cut to the remaining galaxies to construct a high-purity Radio-Likely Prior list. Because stellar mass correlates strongly with star formation rate, this step effectively isolates the population most likely to emit detectable radio continuum.
3. **Map Masking:** Use the MIGHTEE DR1 release catalog to identify and mask bright, complex radio sources (e.g., extended AGN jets $> 50\,\mu$Jy) that violate the point-source assumption of the deblender.
4. **Bayesian Deblending:** Run the `XID+` framework (using the `Stan` probabilistic programming language) to sample the full posterior probability distribution of flux densities for all prior sources in the confused MIGHTEE maps.
5. **Scientific Analysis:** Extract the deblended flux catalogue, assess goodness-of-fit (Bayesian p-values), and compute the Euclidean-normalized 1.4 GHz radio source counts (converted from the 1.3 GHz MeerKAT observations for comparison with the literature).

Expected Outcomes:
The primary outcome will be one of the deepest catalogues of faint radio sources in the MIGHTEE XMM-LSS field to date. The student will accurately measure the faint-end of the radio source counts, pushing well below the traditional confusion limit. Furthermore, the student will quantify the effectiveness of the stellar mass-selected prior approach. The project will culminate in a Master's thesis and has strong potential to result in a lead-author publication in a high-impact peer-reviewed journal.

Skills and Tools (Student's Role):
The student will gain highly sought-after skills in big data astronomy, Bayesian statistics, and machine learning. They will learn to manipulate large astronomical catalogues and images, and write/adapt code in Python. Experience will be gained using advanced statistical software (`Stan/XID+`) and machine learning libraries (`scikit-learn`, `UMAP`, `HDBSCAN`). A basic knowledge of Python programming and extragalactic astrophysics is required.