URSI-RCRS 2024, ARIES/GEHU, Bhimtal, India, 22 - 25 October, 2024

An automated radio data analysis pipeline for the GMRT

Narendra Nath Patra[(1)], Abhinav Narayan[(1)], Atharva Mirashi[(1)]
(1) Department of Astronomy, Astrophysics & Space Engineering, Indian Institute of Technology Indore, Indore 453552, India, e-mail: naren@iiti.ac.in

Radio observations are crucial in probing galaxy formation and evolution processes. Specifically, the neutral Hydrogen in galaxies plays a crucial role. It acts as the reservoir of fuel for star formation. Not only that, but it also acts as a powerful tracer of dynamics. However, analyzing radio data and producing science-ready maps and cubes are not trivial. It takes a significant effort to reduce radio interferometric data, specifically at lower frequencies. At these frequencies, radio frequency interferences (RFI) can corrupt the data considerably, and hence, identification and mitigation of those remain challenging to date.

Aiming to address these issues, we have developed an automated pipeline to analyze radio interferometric data. We implement AI/ML-based algorithms to efficiently identify the RFI and other artifacts. As a first step, our AI/ML algorithms are optimized to work with the GMRT data, both narrow-band (GMRT) and wide-band (uGMRT). This pipeline performs automated flagging and calibration of radio data, producing outputs that are comparable to, or even surpass, those achieved through manual analysis techniques.

The automation of data analysis not only improves consistency but also significantly reduces the time and labor required to process large datasets, making it possible to undertake extensive surveys with greater ease. The pipeline typically takes 20-30 mins to process a GSB data of size ~ 10-12 GB in a standard server class machine/workstation. It takes about an hour and a half for GWB data of size ~ 60-80 GB. Currently, our AI algorithms are optimized for uGMRT band-5, band-4, and band-3.

This pipeline has helped us to conduct a large HI survey using the GMRT, the GARCIA survey. Under this survey, we have undertaken to analyze HI data of 529 nearby galaxies and use this data set to probe galaxy formation and evolution process. By automating the analysis process, we can more effectively prepare a large set of HI spectral cubes to investigate the intricate processes involved in galaxy formation and evolution. This advancement has significant implications for future large-scale surveys (both line and continuum) conducted with the uGMRT, ultimately providing the astronomical community with high-quality, processed data for further exploration and discovery.