

# Radio Survey Data Analysis in the Visibility Domain

*Tuesday 18 June 2019 14:55 (15 minutes)*

I will talk about a use case where new scalable Bayesian methods may be used for detecting and characterize galaxies directly from visibilities of large-scale radio continuum surveys (Rivi & Miller 2018, Rivi et al 2019, Malyali et al. 2019). The analysis of radio surveys has traditionally relied on a set of image reconstruction techniques. However the imaging process may introduce artifacts and correlated noise distributions, with subsequent estimates of scientific parameters suffering from systematic errors that are difficult to accurately estimate. Until recently this has not been a major issue, but the increased sensitivities and size of the forthcoming generation of radio interferometers, such as SKA will allow new scientific measurements, such as weak lensing, that require more reliable and complete source catalogues, meaning higher accuracy in galaxy detection and characterization. An alternative approach to image reconstruction is to work directly in the visibility domain, where the data originates and it is not yet affected by the systematics introduced by the imaging process. Modelling of Direction Dependent Effects, obtained in observations of large fields of view or from radio telescopes with non-coplanar baselines, may also be easily introduced in model fitting techniques for galaxy parameter estimation and data calibration. This novel approach is very promising but also computationally very challenging because of the large size of datasets that must be processed and the source number density expected to reach. For example, a nominal weak lensing survey using the first 30% of Band 2 will require 30 kHz frequency channel bandwidth and 0.5 seconds sampling time to make smearing effects tolerable, meaning about 20 PB of raw visibilities per pointing for 1 hour integration time with the current design of SKA-MID. Adding that the expected number of sources for such surveys is in the order of 104 per field of view, it is clear that this analysis will require tools exploiting High Performance Computing (HPC) infrastructures and has to be performed where the data is stored.

**Author:** RIVI, Marzia (Istituto Nazionale di Astrofisica (INAF))

**Presenter:** RIVI, Marzia (Istituto Nazionale di Astrofisica (INAF))

**Session Classification:** Session 3a: From data to science and back: data processing and pipeline management - Chair: G. Polenta