



A fully automated casa-based pipeline for GMRT data analysis

Ishwara Chandra CH⁽¹⁾ and Russ Taylor⁽²⁾

(1) NCRA-TIFR, Pune University Campus, Ganeshkhind, Pune – 411007, India; email: ishwar@ncra.tifr.res.in

(2) IDIA, UCT, Department of Astronomy, Cape Town, 7701, South Africa; email: russ@ast.uct.ac.za

Here we present a fully automated casa based pipeline for legacy and upgraded GMRT data analysis. The pipeline is extensively tested on 52 pointings of ELAIS-N1 field at 610 MHz using legacy GMRT system and also works good for upgraded GMRT. The input to the pipeline is the multi-source FITS file and other book keeping information regarding the calibrators, target name, channels to average and other imaging and calibration parameters. For convenience, the pipeline was divided into two parts as calibration and imaging. The salient features of the pipeline are (i) it is fully casa-based, so no external packages required, (ii) fully automated from end-to-end, from importing the raw data into casa till the final self-calibrated science ready image. Several novel flagging parameters ensures that minimal data is flagged. Initial flagging is being done using casa task flagdata after which delay, bandpass and gain calibration is carried out. Post-calibration, the data was flagged again and re-calibrated. Channel averaging is included to the extent to keep the bandwidth smearing negligible. The imaging will be carried out using the task tclean. Four rounds of phase-only self-cal and then 5 rounds of amplitude and phase self-cal will be carried out. We have observed that the residual based flagging performs better than regular flagging in the self-calibration loop. The solution interval was kept at 8 minutes for the first round and 1 minute for the fourth and final round of phase-only self-calibration. The final solution interval for calibration was limited to 1 minute in the fourth and fifth round. However, there are also some limitations such as (i) the pipeline assumes that a major fraction of the data is clean and (ii) compute intensive. In this talk we discuss the main features and possible directions to scale this to large data.