National Centre for Radio Astrophysics

Internal Technical Report

GMRT/TGC/25Aug2025

TGC Automation: Phase-I Tasks

Yogesh Maan, Santaji Katore, Deepak Bhong, Nilesh Raskar, Jitendra Kodilkar, Sanjay Kudale, C. H. Ishwara Chandra, Yashwant Gupta, Navnath J. Shinde and Manthan S. Thombare

Email: ymaan@ncra.tifr.res.in

Objective: To provide a detailed background for the TGC automation requirements, list the identified tasks to be automated, and provide detailed accounts of primarily two tasks: (1) Automated decision on phasing quality using in-memory phasing, and (2) Automated decision on disk selection for the beamformer observations at GMRT.

Revision	Date	Modification/ Change
Ver. 2	25 Aug. 2025	Final Version

TGC Automation: Phase-I Tasks

Yogesh Maan, Santaji Katore, Deepak Bhong, Nilesh Raskar,
Jitendra Kodilkar, Sanjay Kudale, C. H. Ishwara Chandra,
Yashwant Gupta, Navnath J. Shinde and Manthan S. Thombare

1 Introduction

The ability to run an end-to-end observation with no or minimal human intervention would minimize the scope for human errors, bring objectivity to subjective decisions that need to be taken by the operators (and in some situations even the users) at various stages, as well as save some precious telescope time (and instead use that in collecting useful data). To achieve such an ability, automation of tasks at various stages of a typical observation is considered.

Considering a reasonable automation of the regular GTAC observations, various steps involved in a typical observation are primarily divided into three parts: (1) Observation setup, (2) Conducting the observation, and (3) Concluding tasks (e.g., preparing and sending the observation log to the users). These steps basically constitute various components, automating which is either necessary or desirable, to achieve running a full observation with no or minimal human intervention. It was suggested to streamline this automation process in three phases, where the first phase involves mostly the tasks that can be automated with least complexity and/or the tasks for which automation would benefit the users significantly. Second phase involves the tasks which are still crucial to automate for a reasonable automation of the observations but it might require slightly longer time to implement the required changes. The third phase involves the tasks which might require significant changes in the system elsewhere, and, hence might involve correspondingly significant contributions from different teams and be realizable only over a long term.

2 Phase-I Automation Tasks

There are primarily three tasks under the Phase-I automation which are listed below with brief description.

1. *Automated decision making on the quality of phasing for beam-mode observations, and related developments.*
   The beam-mode observations require the signal from a number of antennae in one or more sub-arrays to be coherently added to achieve the full sensitivity. The process of coherent addition requires the phase differences between signals from different antennae to be estimated (with respect to a reference antenna) and then corrected for, i.e., phasing of the array. The quality is often judged by visualizing the phases of signals from various antennae in a plot by the operators, and as a result, the decision whether the phasing is acceptable or not is subjective. In this task, this process of making decision on the quality of phasing of the array is objectified and automated. More details on this task are provided in Section 3.

2. *Automated decision on choosing the disk for beamformer data recording.* The planning for beam-mode observations requires choosing the host machine as well as a particular disk (from several options per host machine) where the data would be recorded. The selection requires the user to carefully estimate the data requirements as well as space availability, however, the latter could change by the time the actual observation starts. This task deals with this selection automatically. More details on this taks are provided in Section 4.

3. *Automated observation log/report.* At the end of each observing session, the operators provide the user a log of the observations, which includes various system parameters, location of the recorded data files as well as a summary of any issues or problems faced during the session. This task focuses on automating the generation of this observation log or report, including automatically fetching various parameters and data locations from the relevant parts of the system. More details on this task will be provided in the form of a separate internal report once the task is concluded.

The recently developed command-file generator (UI as well as its backend; Bhong et al., 2023; Katore et al., 2025), which is also being continuously upgraded to be compatible with newly released observing modes, is another major task towards automation. However, the command-file generator related work is summarized separately, and hence, is not a part of this report.

3 Automated decision on phasing quality and in-memory phasing

3.1 General background

In the beamformer mode of observations at GMRT, the antennas in a sub-array are phased, i.e., the signals from different antennas are added in-phase, to achieve the cumulative sensitivity of all the antennas in the sub-array. The gain of a phased-array can be given by a slightly modified form of Ruze's equation (Ruze, 1966):

$$G = G_0 e^{-(2\pi\phi_{rms})^2}, \tag{1}$$

where, G_0 is the gain for ideal phasing, i.e., the signal from the individual elements or dishes are added fully coherently, and G is the actual gain achieved for a case where the signals from different dishes are added with an RMS error of ϕ_{rms} in their phases (*in units of* 2π). The quality of phasing can be checked by estimating the relative gain, G/G_0, which hereafter, we refer to as the relative sensitivity. For a given set of phase-errors from different elements or dishes, the relative sensitivity can be estimated quantitatively by assuming the incoming electromagnetic wave at each of the dishes to be of the same amplitude, and then computing the vector-sum, folding in the phase-errors. As the gain is decided by the power of the phased-array signal, the square of the ratio of the resulting amplitude to the originally amplitude provides the relative sensitivity. A demonstration is provided in Figure 1 — estimates of relative sensitivity from simulations of phase-errors uniformly distributed over a number of pre-decided phase-ranges are shown as points, along with the prediction from the above equation 1. Note that the RMS of phase-errors distributed uniformly and randomly between ϕ_{min} and ϕ_{max} is given by $(\phi_{max} - \phi_{min})/\sqrt{12}$.

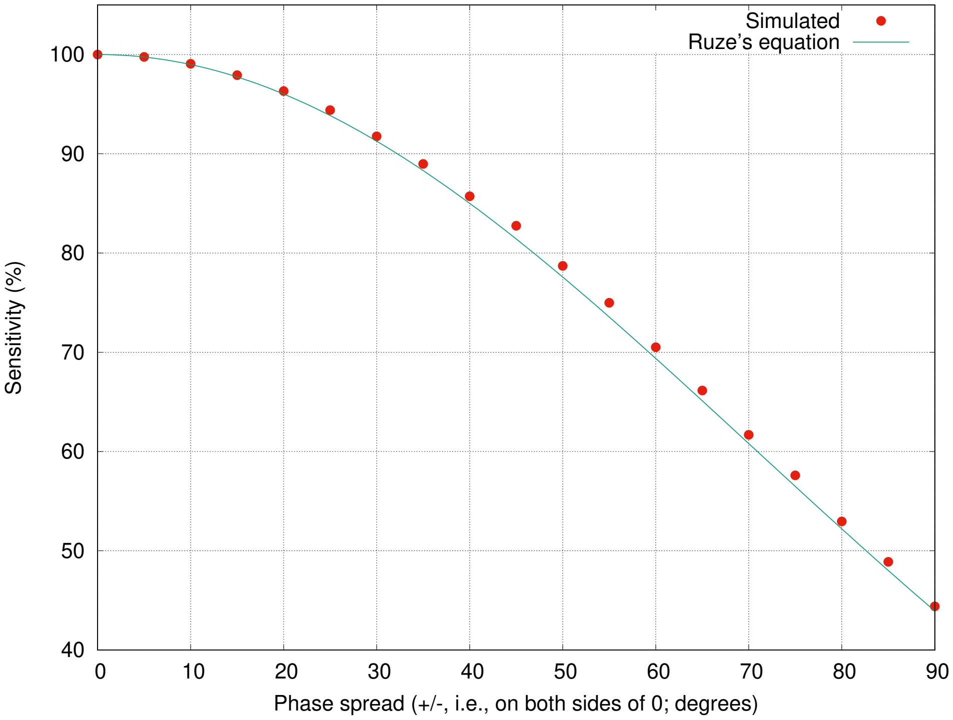


Figure 1: The percentage achievable relative sensitivity as a function of the spread in the phase-errors. The points show the estimates from simulations while the continuous line shows the predictions from Equation 1.

3.2 Decision making on phasing-quality

Phasing the sub-array of a number of antennas at GMRT constitutes observing a phase-calibrator, estimating the phase-offsets for each of the antennas at each of the frequency channels with respect to a reference antenna, and then correcting for these measured phase-offsets in the correlator. Once these phase-offsets are measured and corrected for, any subsequent signals obtained from various antennas are expected to arrive in-phase, i.e., the phases of signals from all the antennas at all the frequency channels are expected to be 0 with respect to the reference antenna. However, various factors, e.g., measurement errors, accuracy with which the phase-offsets could be measured on a specified phase-calibrator, varying ionospheric conditions, radio frequency interference (RFI) in some parts of the band, etc., introduce phase-errors in the otherwise supposedly in-phase signals. As discussed in the previous subsection, these phase-errors reduce the achievable sensitivity.

Achieved sensitivity with respect to the sensitivity in an ideal case (i.e., all the phase-errors were to be zero), is termed as relative sensitivity, S_r, in the remainder of this document.

At GMRT, a specified set of antennas are phased separately for two polarizations. For each of the polarizations, the estimated phase-offsets are estimated and corrected for using methods like `rantsol` or `extract`. After the corrections are made, the residual phase-errors are visualized for various antennas together, and as a function of frequency channels. An example of this visualization is shown in the upper panel of Figure 2. To quantitatively estimate the relative sensitivity, we first analyze the phase-errors and identify the frequency channels where the phase-errors are statistical outliers with respect to the rest of the channels. These frequency channels are then considered as RFI-contaminated, and excluded from the estimation of the relative sensitivity. Examples of such RFI-contaminated channels can be seen as gray-shaded regions in the both the panels of Figures 2 and 3. At each of the remaining channels, the relative sensitivity is estimated using the phase-errors corresponding to each of the antennas, as discussed in the previous subsection. The relative sensitivity thus computed is shown as a function of frequency channels in the lower panels in Figures 2 and 3. Note that the relative sensitivity shows large fluctuations at some of the channels, and such channels are close to or in between RFI-contaminated channels (gray-shaded regions). These are primarily the channels which are contaminated by RFI but escaped their identification solely based on the phase-error data (e.g., these might have been found as outliers if the amplitudes or intensities were also available). Nevertheless such channels only a small fraction of total number of channels. The *net relative sensitivity* across the band is estimated as median of the relative sensitivities at all the channels, which is again immune to the above fluctuations as long as the remaining RFI-contaminated channels are less than 50%. Figures 2 and 3 shows the phase-errors and corresponding estimates of the relative sensitivity at all the channels as well as the net relative sensitivity over the full band for the two polarization channels, respectively.

3.3 Implementation details and In-memory phasing

The traditional phasing methods (using rantsol/xtract) typically have the following steps.

1. Start a scan on a calibrator source suitable for the longest and shortest baselines of antennas in the sub-array.

2. Record data for small amount (typically 40 s or so) and calculate phase offsets wrt to a reference antenna using tools like rantsol/xtract.

3. Apply the phase offsets at the designated set point in the signal chain, i.e., to the correlator.

4. Restart the scan.

5. Record data again and manually check the residual phase-errors.

6. If the phase-errors are close to zero: phasing is successful, exit procedure, otherwise: repeat the above steps.

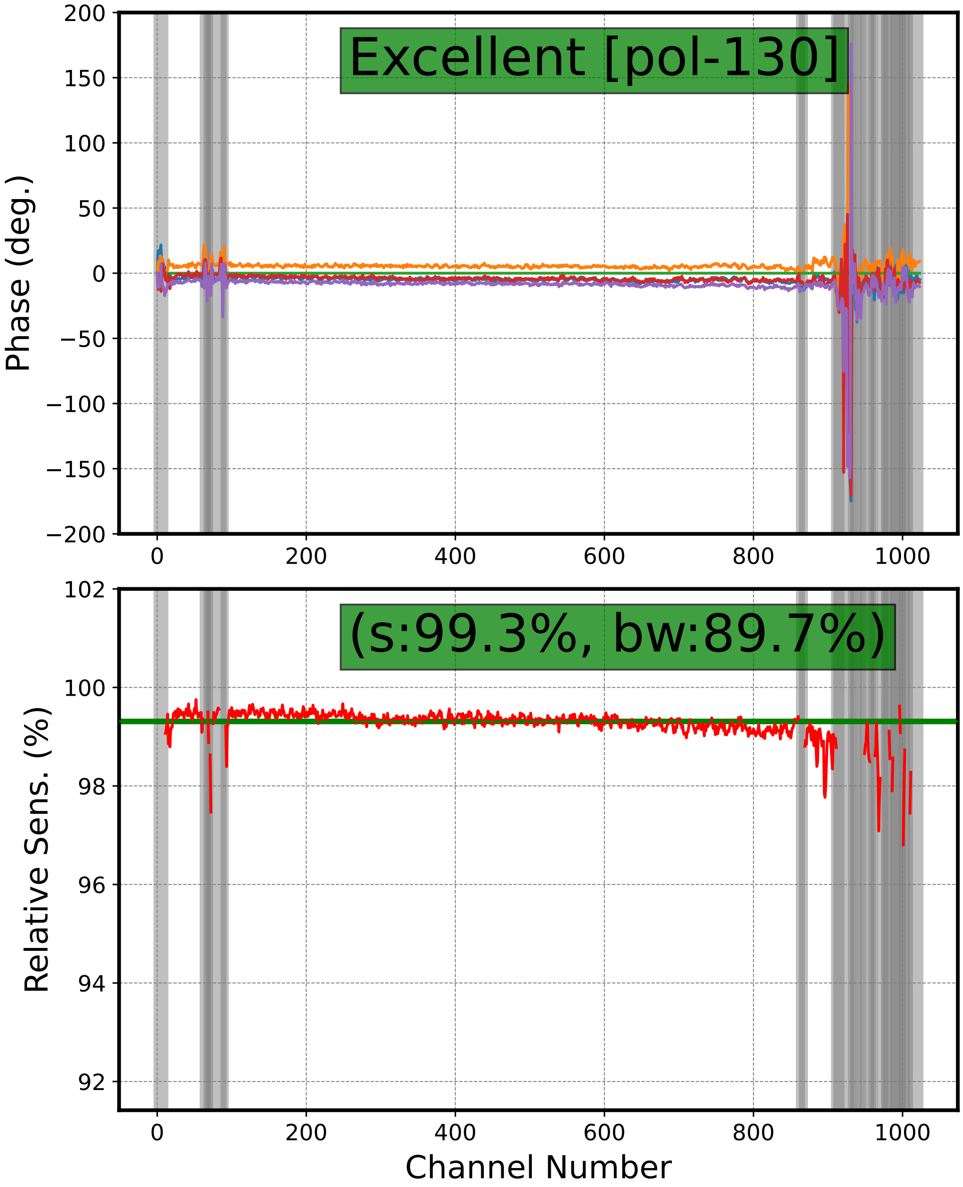


Figure 2: The upper panel shows the residual phase-errors for various antennas after phasing in different colors and as a function of frequency channels. The lower panel shows the estimated relative sensitivity as a function of frequency channels in red color, as well as the median relative sensitivity as a horizontal green line. The green shaded legend in the upper panel shows a qualitative decision on the phasing quality (in this case “excellent”) and polarization identifier (“pol-130”), while that in the lower panels shows the quantitative measure of the net relative sensitivity (in percentage), and the percentage bandwidth that was found to be usable (i.e., not RFI-contaminated). The gray-shaded regions in both the panels mark the channels which are identified as RFI-contaminated.

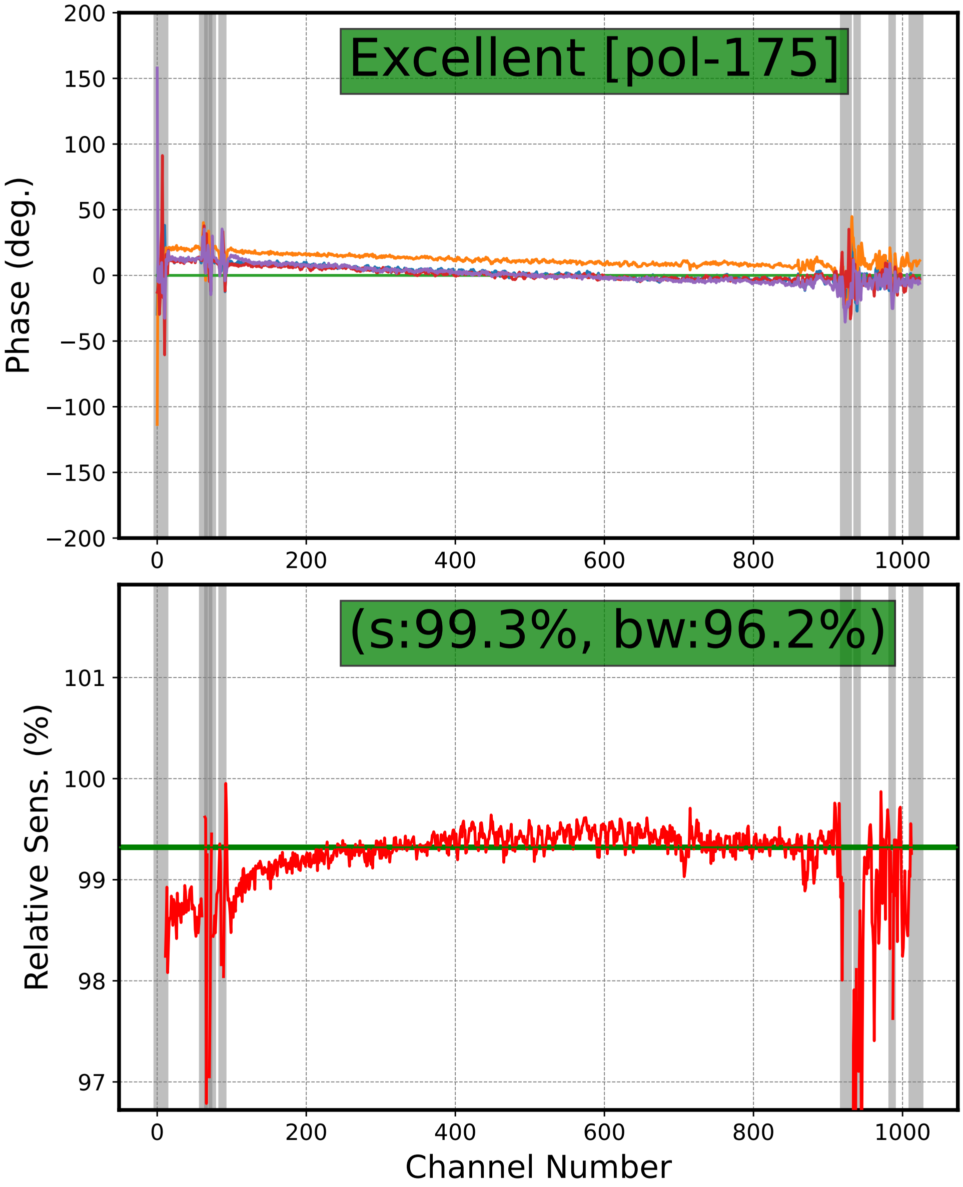
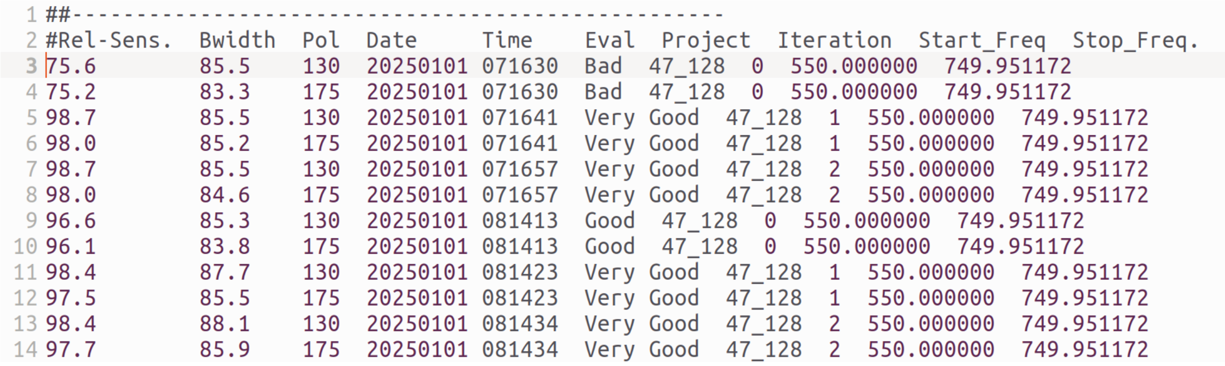


Figure 3: Same as in Figure 2 but for the second polarization channel.

Once phased, the sub-array remains in phase as long as ionospheric conditions remain stable.

In the above traditional methods, the phasing requires a full scan-restart loop (first scan to estimate the phase-offsets and then the second scan to verify the phasing), which adds delay. Moreover, if the phasing quality is not satisfactory, the above added delay grows rapidly with every phasing iteration. To eliminate this overhead, a new method called *in-memory phasing* is developed and implemented. Here, instead of writing data to a file and then reading from the file to estimate the phase offsets, phases obtained from the visibility-data are temporarily stored in shared-memory, and the phase-offsets are estimated using these data over a specified integration time, while the scan is still ongoing. Once the phasing offsets are estimated, these are *virtually* applied to the data from the same ongoing scan in the subsequent integrations and the quality of the phasing is estimated using the methodology discussed in the previous subsection. If the phasing quality is above a specified threshold, then the scan is stopped and the phase-offsets are the *actually* applied to the correlator. If the phasing quality is not satisfactory, then the next iteration of estimating the phase-offsets and applying those *virtually* is conducted using the new data on the source in the same ongoing scan. This process is repeated until the phasing quality reaches above a specified threshold or the maximum number of iterations are done, whichever happens earlier. As this method avoids a full scan-restart for each of the iterations, a lot of the overhead delays are saved.



```
##---------------------------------------------------
#Rel-Sens.  Bwidth  Pol  Date     Time    Eval  Project  Iteration  Start_Freq  Stop_Freq.
75.6        85.5    130  20250101 071630  Bad  47_128  0  550.000000  749.951172
75.2        83.3    175  20250101 071630  Bad  47_128  0  550.000000  749.951172
98.7        85.5    130  20250101 071641  Very Good  47_128  1  550.000000  749.951172
98.0        85.2    175  20250101 071641  Very Good  47_128  1  550.000000  749.951172
98.7        85.5    130  20250101 071657  Very Good  47_128  2  550.000000  749.951172
98.0        84.6    175  20250101 071657  Very Good  47_128  2  550.000000  749.951172
96.6        85.3    130  20250101 081413  Good  47_128  0  550.000000  749.951172
96.1        83.8    175  20250101 081413  Good  47_128  0  550.000000  749.951172
98.4        87.7    130  20250101 081423  Very Good  47_128  1  550.000000  749.951172
97.5        85.5    175  20250101 081423  Very Good  47_128  1  550.000000  749.951172
98.4        88.1    130  20250101 081434  Very Good  47_128  2  550.000000  749.951172
97.7        85.9    175  20250101 081434  Very Good  47_128  2  550.000000  749.951172
```

Figure 4: An example of the summary output that is logged after every phasing iteration.

The phase-errors from the in-memory phasing are used to make a decision on the quality of phasing, by estimating the relative sensitivity, as discussed in the previous subsection. The estimated relative sensitivity for each iteration of phasing, for every project that uses the in-memory phasing, is logged into a pre-decided file as one row. The other parameters that are logged in along with the net sensitivity are: usable bandwidth (estimated from the phase-error data as discussed in the earlier section), polarization channel (conventional denotations of 130 or 175), date and time (in YYYYMMDD and HHMMSS formats), a qualitative evaluation of the phasing-quality (one of the "Excellent", "Very-Good", "Good", "Okay", and "Not-Okay" flags), project code, iteration number, and start and end frequencies (in MHz) of the band for which phasing is being done. An example of such logged in data from two phasing rounds (and corresponding phasing iterations) is shown in Figure 4.

Every phasing iteration also results in making and saving plots for both the polarization channels similar to those shown in Figures 2 and 3. For posterity and any long term uses, the phase-errors from every iteration are also saved as `numpy`'s `npz` format. All these logged data will be available at GMRT for all the projects which use in-memory phasing.

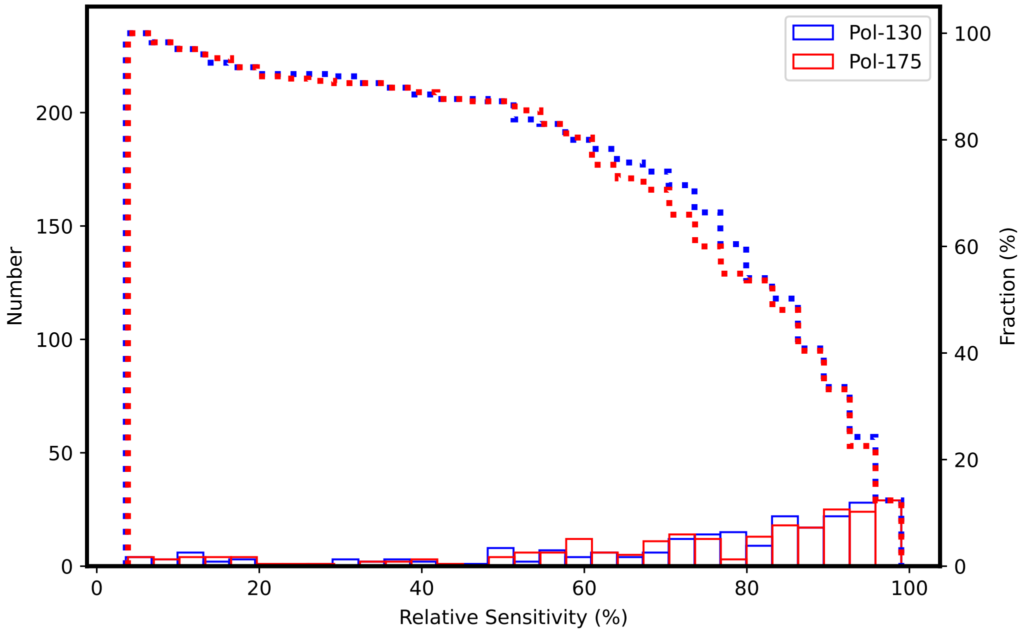


Figure 5: Band-3 statistics of the achieved phasing quality: The bar plots show the histograms of the achieved relative sensitivities in the final phasing rounds of all the observations which have used the in-memory phasing so far, at Band-3. The two colors correspond to the two polarization channels, as indicated by the legends. The cumulative histograms for the two channels are also shown in the same colors but with dotted lines. *Many of the data points with less than and around 50−70% relative sensitivity in this plot have resulted from some of the initial test observations as well as observations that were heavily affected by ionospheric scintillation.*

3.4 Automation requirements

As mentioned in the previous subsection, currently the in-memory phasing is repeated until the phasing quality reaches above a specified threshold or the maximum number of iterations are done, whichever happens earlier. The specified threshold, however, needs to be chosen realistically, considering various factors that might depend on the observing frequency and other aspects. For example, RFI is often detrimental to phasing at Band-2. Furthermore, the achievable phasing-quality is often much higher at the higher frequencies than that at lower frequencies, due to the rapid ionospheric variations at lower radio frequencies.

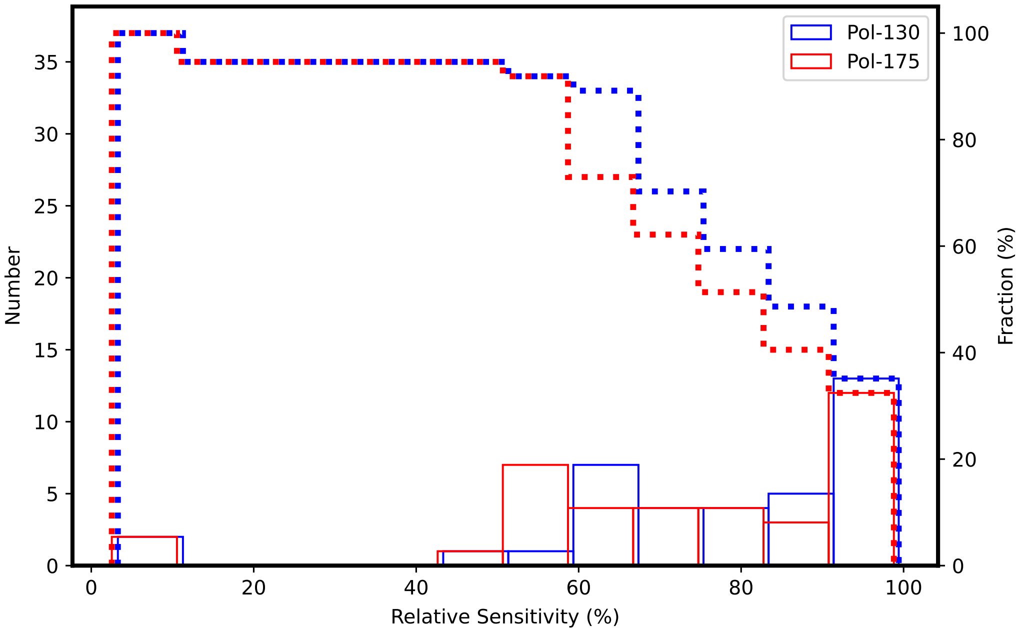


Figure 6: Band-4 statistics of the achieved phasing quality: The details for various histograms are same as those in Figure 5, but now for Band-4. *Many of the data points with less than and around 50−70% relative sensitivity in this plot have resulted from some of the initial test observations as well as observations that were heavily affected by ionospheric scintillation.*

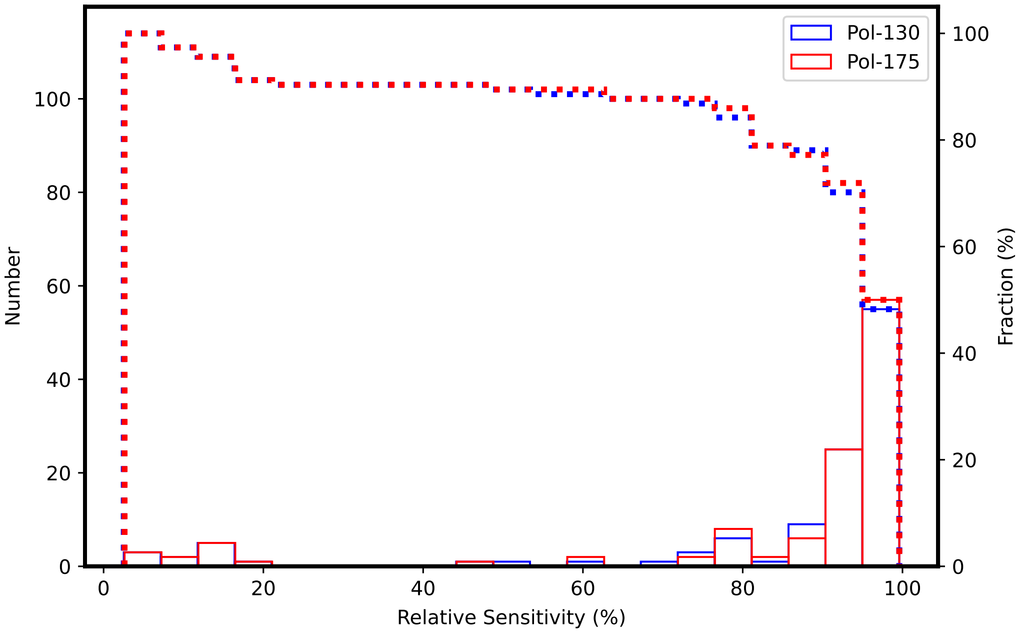


Figure 7: Band-5 statistics of the achieved phasing quality: The details for various histograms are same as those in Figure 5, but now for Band-5. *Many of the data points with less than and around 50−70% relative sensitivity in this plot have resulted from some of the initial test observations as well as observations that were heavily affected by ionospheric scintillation.*

One data-driven way to estimate reasonable thresholds on the phasing quality at various frequency bands is to see the statistical trends from the data that has been logged in from various projects which have already used the in-memory phasing. Using the data that we have so far at different frequency bands, the histograms of the final achieved relative sensitivity in each of the phasing rounds (i.e., that from the last iteration) are shown in Figures 5, 6 and 7. Note that these histograms include data from some of the initial tests as well as those from the observations which failed due to heavy ionospheric scintillations. Nevertheless, using these limited data so far, we see that 80% of the observations at Bands-3, 4 and 5, achieve relative sensitivities of about 65%, 65% and 90%, respectively. Again, these initial trends are currently limited by the factors mentioned above and more data in future will provide the reasonable trends which then can be used by the observatory to specify the reasonable achievable relative sensitivities at various bands of GMRT.

4 Automating the selection of disk for beam-data recording

The beam-data at GMRT are recorded in disks mounted at four different host machines: `gwbh7, gwbh8, gwbh9` and `gwbh10`. The host machines are mapped to beam-numbers, and it gets automatically selected as soon as the user selects a particular beam. Each of these host machines have eight disks of different capacities mounted on them. At the time of preparing the observation setup files, the user needs to choose the disk on which the data should be recorded. Often the user chooses the disks which have most space available. However, it often becomes an issue, as the disks keep getting filled up uniformly and then a long observation which might require large amount of space (e.g., more than a TB) can not be recorded on a single disk. To circumvent this issue, an automatic directory selection for the GWB beam data recording has been developed and implemented.

The automatic directory selection is implemented at two stages: command file conversion stage and the command file runtime stage. During the command file conversion (from the setup sent by the user), the required disk space is estimated on the basis of inputs from observation setup. The required disk space is then cross-checked against the available disk space on the specified beam recording host machine. This information is stored in a file with the respective project in JSON format. A few more details of the algorithms used at the two stages are provided below.

First Stage (command-file generation) check

Following algorithm is used for selection of recording disks at the command file conversion stage.

- Find if a single disk with the required space is available for recording of data for the full observation.
- If no such single disk is available then choose multiple disks.
- If required disk space is not available even cumulatively from multiple disks on the required host machine, then give warning at command file generator stage.

All the beam configuration parameters along with the above directory information are stored in a JSON file on a TGC server machine for future use, at the time of running the actual command file.

Second Stage (during the command file execution) check

At this stage, just before issuing the beam recording command, the name of the earlier identified disk is picked from the JSON file, and checked for availability of required space for the observation. If the command fails to get the required space, an audio alarm is generated.

The automated disk selection is being used for all the beam data recordings since January-February 2025.

5 Phase-II Automation Tasks

The automation tasks that will be taken up as a part-II of the TGC automation were identified after extensive discussion in the Tel-Ops meetings. The identified tasks, along with their priorities are briefly listed below.

1. Automatically check and re-iterate power-equalization, i.e., the counts are brought within a small allowance around a specified threshold (priority-1).

2. Fix the issue with ”goto src” (sometimes the command gets wrong target location) and make a provision to bypass the command (priority-1).

3. Integrate the Online-RFI Filter settings in the command file generator (priority-1).

4. Automatic submission of setup/command files via the command-File generator (priority-2).

5. Validate the source visibility, i.e., check the source rise/set time, at the time of command-file generation stage (priority-2).

6. Develop a provision to interrupt only the current scan, instead of stopping the whole command file (priority-2).

7. Feed rotation related, fix the issue with occasional command failure and the 2 minutes long waiting time, and automatically resend the command only for the required antennae (priority-2).

8. Related to automatically reading and checking the return-status of commands: (a) start of pulsar DAS chain, (b) start of LTA file recording, (c) start of Beam data recording (priority-3).

9. Related to the Antennae Setup (RF + LO + IF) + GAB setup + Correlator setup (RFI filter + Walsh + Pulsar mode): A mechanism to automatically check correlation matrix and raise alarm accordingly (priority-3).

6 Summary

We have presented a summary of the automation tasks that are completed or being conducted as a part of Phase-I of TGC automation. In particular, we have provided a detailed account of the work on automatically determining the phasing quality using in-memory phasing and the automated decision on disk allocation. Towards the end, we also provided a list of tasks to be taken up in Phase-II of TGC automation.

References

Bhong, D., et al. 2023, NCRA Internal Reports, R317

Katore, S., et al. 2025, NCRA Internal Reports, Under-Preparation

Ruze, J. 1966, Proceedings of the IEEE, 54, 633