Progress Report On Continuum Survey Via KAT-7

Ermias A Kassaye

April, 2015

Overview Of The research

The candidate, Ermias K., studied Active Galactic Nuclei (AGN) observed by the KAT-7 telescope that are likely to be good calibrator sources. For the research, 48 potential calibrators were identified from various surveys or catalogues such as NVSS, SUMSS, etc. Majority of the selected sources ($\sim 90\%$) reside over the southern hemisphere. The sources belong to different AGN groups: about 46% are quasars; $\sim 50\%$ are faint galaxies and two sources are BL Lac objects. The classification based on their spectra indices: $\sim 41\%$ and $\sim 43\%$ have flat and steep spectra indices respectively. Moreover, inverted spectra ($\sim 8\%$) and Giga Hertz Spectrum ($\sim 8\%$) sources are identified in the sample.

Using the sources, the research has achieved the following:

- The position accuracy of the KAT-7 telescope was determined. According to the result, the KAT-7 could measure the position (right ascension and declination) of a source at an accuracy ~ 5 arcsec.

- It was checked whether the KAT-7 is on the same flux density scale with other radio data by using standard flux-density calibrator sources (i.e. 3C147, 3C286, and 3C48). The result showed that the KAT-7 is consistent with other radio data at the level of $\sim 2\%$.

- There is always a systematic error during the process of transferring (bootstrapping) flux-density from the flux-density calibrator to the target source. The systematic error was estimated using the well-studied source (3C123), which has been shown to have a very constant flux density over time. The analysis showed that there is a $\sim 5\%$ systematic error during flux-density bootstrapping.

- Since some of the sources were observed at different days between Oct 2012 and Feb 2013, their flux-density variability with time were statistically analysed. The result is a preliminary indicator whether we would use the sources as flux-density calibrators for future surveys.

The current research uses the above results as an input for improving calibration and source extraction. The initial phase of the research are focusing on extracting sources located within the KAT-7 field of view ($\sim 0.8^{\circ}$). In this case, each potential calibrator would serve as a phase center. To implement this, we initially utilize PKS J1939-6342 for the development of a data reduction pipeline.

After successfully implementing the pipeline to this source, we use it for the other sources in the list. Accordingly, the project goal will be :

- preparing a catalogue to be used as a sky model for calibration as well as utilizing it during the commissioning phase of the MeerKAT.
- getting a clear picture of the intrinsic and extrinsic variability of the sources.
- examining the number of sources as a function of the flux density to identify AGN and starburst galaxies.

  We can also deal with other issues that will come unexpectedly through the research process.

Summary of Progress

So far the research candidate has been reading different publications supporting the research and writing scripts that will be compatible with the softwares (meqtrees, CASA, ...). For this purpose, Pyxis was created by O.M Smirnov to simplify and facilitate the data reduction process. Hence, the candidate has been learning about the Pyxis as well as Meqtrees to effectively utilize them for the aforementioned objectives of the research.

The general steps followed by the candidate to extract of the sources.

- PKS J1939-6342 was selected as a tester for the development of the pipeline. The fact that its stable flux density and compact structure, as confirmed by various radio telescopes and interferometers, made the source to be our first choice for the purpose. Moreover, since it is southern hemisphere source, we could have 8 measurement sets (MS) during the observation periods. This helps us investigating the problem with each MS during the process of developing the pipeline.

- After having the position of the source from SUMSS catalogue and its flux density at 1.83 GHz from the candidate's previous research work, the sky model was prepared.
- After performing self calibration in meqtrees, we obtained the map representing the model minus the source. From this map, we could able to identify the sources using the source extractor software (pybdsm) within the KAT-7 FOV.

The pipeline has been constructed to firstly calibrate each MS and then concatenate them to yield one measurement set. By subtracting the sky model from the calibrated concatenated MS, we extract the source in the FOV. Using the corrected residual map, we implement pybdsm to extract the sources at threshold pixel of 5 and threshold island of 3. By doing so, 5 sources have been identified ranging from 25.5 mJy to 66.8 mJy.

Detail of the pipeline

The pyxis has special functions known as superglobals (i.e., visible to all functions in a particular module). These are the measurement set (MS), the local sky model (LSM), log file (LOG) , and so on. Pyxis consists python functions or known as Pyxis recipes named by "pyxis-*.py". There are also python-readable files (configuration files) assigned by "pyxis-*.conf". These files are used to specify variables necessary for the data reduction purpose to work with "pyxis-*.py". These files are useful to define some specific parameters, rather than putting all the variables in the recipes. All options necessary to simulate in the meqtrees are saved in a profile named by tdlconf.profiles.

It is important to have a separate subdirectory containing pyxis-*.py, pyxis-*.conf, tdlconf.profiles, and the sky model file (appropriatename.lsm.html or appropriatename.txt). We might have received error messages if we put the measurement datasets with the mentioned files in the same directory. Hence, it is preferable to use different directories for MSs and the recipes.

It should be noted that we need to simulate one of the dataset by using meqbrowser (i.e. writing meqbrowser on console will guide us to meqserver and click start) before running the script. The options necessary for the calibration procedure have to be ticked in the TDL compile-time option. The options should also be saved by clicking at the bottom of the box. In this case, all options will be written on tdlconf.profiles and the given file name will appear on the header. The file name will be called by the script when we execute pyxis-*.py for the calibration process.

The tile size option in TDL is set to be 1 by default. If we compile your script with this value, we will encounter an error. So, we need to assign the tile size to be 64 or 128.

Let us consider the pipeline for our purpose. The file containing the recipe is pyxis-kat7.py. We imported all modules necessary to perform the intended task. Some of these modules are:

- Pyxis: The backbone for the entire procedure; it imports all the interfaced tasks
- ms: the task useful to perform MS related actions. For example, some of the MS tasks are:

  ```
  pyxis msfile ms.summary
  pyxis msfile ms.plot_uvcov
  pyxis msfile ms.plotms
  pyxis msfile ms.aoflaggercasa
  pyxis msfile ms.verify_antpos
  pyxis msfile ms.save_tarball
  pyxis msfile ms.listcols
  pyxis msfile1 msfile2  ms.merge[output='msfile_output.MS']
  ```

- lsm: task related to source extraction purpose. For instance

  ```
  pyxis lsm.pybdsm_search[image.fits,output=model.lsm.html,threshold=7]
  ```

- im: module important for imaging purpose. For example,

  ```
  pyxis msfile im.lwimager.make_image[cellsize=30arcsec,npix=512,
  restore=True,psf=True,column=CORRECTED_DATA]
  pyxis msfile im.casa.make_image[cellsize=30arcsec,npix=512,
  restore = True,psf=True,column=CORRECTED_DATA ]
  ```

- stefcal: interface to for calibration related task.

The first line of the code put all parameters for the calibration.
Producing a model form the raw data.

```
def calibrate(msname='$MS', lsmname='$LSM',tdlsec='$CALSEC',
              column='$COLUMN', do_dE=False, args=[],**kw):
```

where msname is measurement set, lsmname is the local sky model, tdlsec is the options from the TDL, column is the column that we need to be analysed (i.e DATA, MODEL, ...)

The following line of code will do the selmfcalibration be getting the options defined above.

```
stefcal.stefcal(msname,section=tdlsec,options=options,args=args)
```

The calibration of each dataset is defined using the initial sky model (lsm0).

```
def cal_ms(lsm0='$LSM0', start=0, stop=4):
```

This line of code calibrates one after the other using the initial sky model (lsm0) for all dataset under the list.

```
run_cmd = lambda : calibrate(lsmname=lsm0)
pper("MS",run_cmd)
```

After calibrating each dataset, it will concatenate and produce one measurement set. Using the concatenated MS, images will be produced with the following line of codes.

```
ms.virtconcat(output=CONCAT_MS)
v.MS = CONCAT_MS
im.make_image(restore=True, psf=True, restore_lsm=False)
```

Using the concatenated corrected residual image(Data - Model), the sources around the phase center will be extracted. The following lines perform the procedure. The third line will append the intial sky model with the catalogue of the extracted sources from the corrected residual image.

```
lsm.pybdsm_search(thresh_pix=5 , thresh_isl=3)
v.LSM = lsm.PYBDSM_OUTPUT
x.sh("tigger-convert --append $LSM $LSM0 $LSMFINAL -f")
```

We use the following command in the console to run the script:

```
pyxis -j8 OUTDIR=Test cal_ms
```

where $-j8$ corresponds to the number of ms that will be excited. If we have 5 ms in the directory, we use:

```
pyxis -j5 OUTDIR=Test cal_ms
```

We can assign different directory name for the output in the option OUTDIR (eg. OUTDIR=PKSJ025-2603, ...). The user will assign a directory name according to his/her preference.

Cross Validation

After extracting the sources, we cross-check their positions with their corresponding catalogue positions. This will serve as a means to distinguish whether the detected sources are real or fake. Secondly, we predict the flux densities of the sources at our observing frequency and compare them with the value obtained from our source extraction analysis.

We can estimate the flux density, S_{kat}, of a source at the KAT-7 observing frequency, ν_{kat}, using:

$$S_{kat} = S_{cat} \left(\frac{\nu_{cat}}{\nu_{kat}} \right)^{\alpha}, \tag{1}$$

where S_{cat}, ν_{cat} and α are flux density of a source obtained from the catalogue, the frequency of the survey in which the source was observed, and spectral index, respectively.

Since the spectral indices of the extracted sources are not available in the catalogues, we fairly use $\alpha \sim$ -0.7. This is due to the fact that $\alpha =$ -0.7 is a typical indicator of the non-thermal radio sources (i.e., detailed explanation is also available in the candidate Msc. thesis, pg 15-20).

We also need to compare the theoretical RMS with that of the images using:

$$\Delta S = \frac{2K_B T_{sys}}{\eta A \sqrt{\Delta\nu \Delta t n(n-1)}}, \tag{2}$$

where Δt is the total integration time. The other values in the above equation are listed in Table 1 shown below.

Plan for the next 6 months

- Apply the pipeline for other calibrator sources in the list.
- Analyse the result and prepare the catalogue
- Write a draft paper for publication

Table 1: The key parameters of the KAT-7.

Parameters	values
Number of antennae (n)	7
Dish diameter (D/m)	12
Dish Area (A/m^2)	110
Aperture efficiency (η_a)	0.66
Central frequency (ν/GHz)	1.83
Bandwidth ($\Delta\nu$/GHz)	0.25
Wavelength (λ/cm)	16.41
Minimum baseline ($b_{\mathtt{min}}$/m)	26
Maximum baseline ($b_{\mathtt{max}}$/m)	185
System temperature (T_{sys}/K)	$\leq$ 35 across the entire frequency band (∽30 average) for all elevation angles
Synthesised resolution ($\theta_{\mathtt{res}}$)	$\sim 4'$
Primary beam (θ_{pb})	$\sim 54'$
Sky coverage (deg)	-80 to 50 (Dec) and 0 to 360 (RA)

Source confusion

In flux limited survey, only sources above a given 'cut off' flux density limit, S_c, will be counted while fainter sources will discarded. Given a finite beam size, many faint sources could crowd into a single beam area, since the distribution of sources is random. Then, it is possible that the total flux density is above the cut-off and this collection of weaker sources will be erroneously counted as a single source with $S > S_c$. This is referred to as Source Confusion; this is caused by a combination of high sensitivity and insufficient angular resolution. This situation is normally encountered at low frequencies. This analysis was used to obtain a dependence of the number of sources as a function of source flux density.

The above analysis is not realistic according to Condon because of the following reasons.

- The details of the instrument beam shapes, including sidelobes, must be taken into account.
- The cutoff in a survey is the recorded deflection, D_c, not the source flux density cutoff S_c.

Extragalactic Continuum Source Confusion

The total number of sources with flux densities between S and $S{+}dS$, dn, is

$$dn = \int n(S) d\Omega dS \tag{3}$$

If the response of the is

$$x = fS \tag{4}$$

where $f(\theta, \phi)$ is the normalized antenna pattern. Then

$$dx = fdS \tag{5}$$

The standard expression for $n(s)$ is

$$n(S) = kS^{-\gamma} \tag{6}$$

where k is a constant expressing the ratio of the number of sources per unit area to cut-off flux density, that is, $k = \gamma N_c S_c^{\gamma}$ where N_c is the number of sources at the cut off flux density S_c and $\gamma > 0$. Using the expressions, the average number is

$$dn = k \int \int [f(\theta, \phi)]^{-\gamma-1} S^{-\gamma} d\Omega dx \tag{7}$$

Another form is

$$dn = k\Omega_e \int x^{\gamma} d\Omega \tag{8}$$

where we have defined the effective beam, Ω_e, as

$$\Omega_e = \int [f(\theta, \phi)]^{-\gamma-1} d\Omega \tag{9}$$

where $f(\theta, \phi)$ is the normalized antenna power pattern. The distribution dn is Poisson. The dispersion of the response from 0 to the cutoff D_c is

$$\sigma^2 = \int_0^{D_c} x^2 dn \tag{10}$$

when Eq.8 and the expression for x are substituted into Eq.10, one obtains

$$\sigma^2 = \left(\frac{k\Omega_e}{3-\gamma} \right)^{\frac{1}{\gamma-1}} D_c^{3-\gamma} \tag{11}$$

For $2 < \gamma < 3$. Expressing D_c by a factor q times σ, we have

$$\sigma^2 = \left(\frac{k\Omega_e}{3-\gamma}\right)^{\frac{1}{\gamma-1}} (k\Omega_e)^{\frac{1}{\gamma-1}} \tag{12}$$

The first factor in the above equation depends only on the number of count exponent and the choice of q. The second factor is related to the angular resolution as given by Eq.9. Thus the signal-to-noise ratio of the faintest sources observed with a given instrument depends on two parameters.

- The average number of sources in the beam, ΩN_c at the limiting flux S_c, and
- The power n of the source distribution density function.

Therefore, the highest the angular resolution, the deeper the survey can go before confusion occurs. Usually there is a uniform flux density cutoff limit S_c. This flux density is a confusion limit for the telescope at the chosen frequency. A q value of 5 is usually considered to be acceptable; this corresponds to a probability of 10^{-6} for erroneous source identification

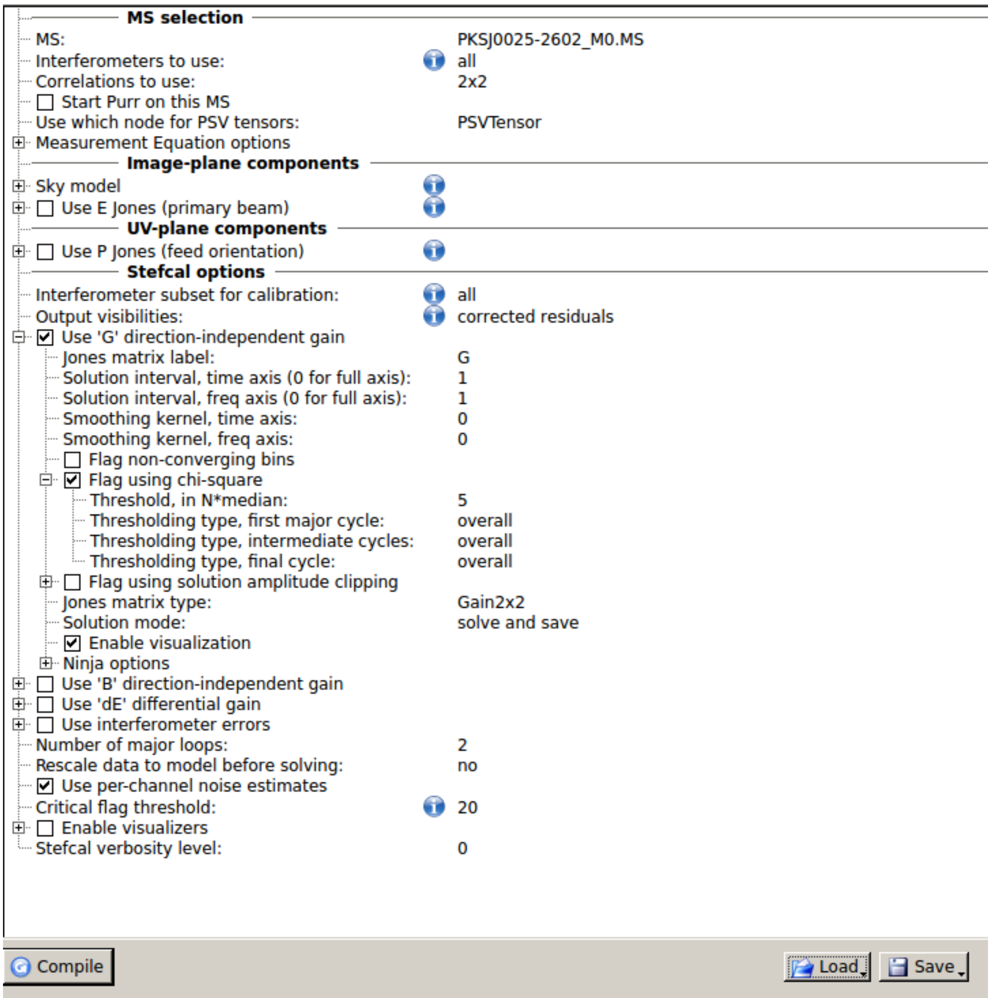

Figure 1: The TDL (Tree Definition Language) for gain (G) calibration options. After saving the options clicking at the bottom, we need to click on the 'compile' button to run the simulation. We need to click on the 'bookmarks' button at the top. The bookmarks have options to look at the input and output visibilities, so we need to click on 'input visibilities: inspector plot' and 'output visibilities: inspector plot'.

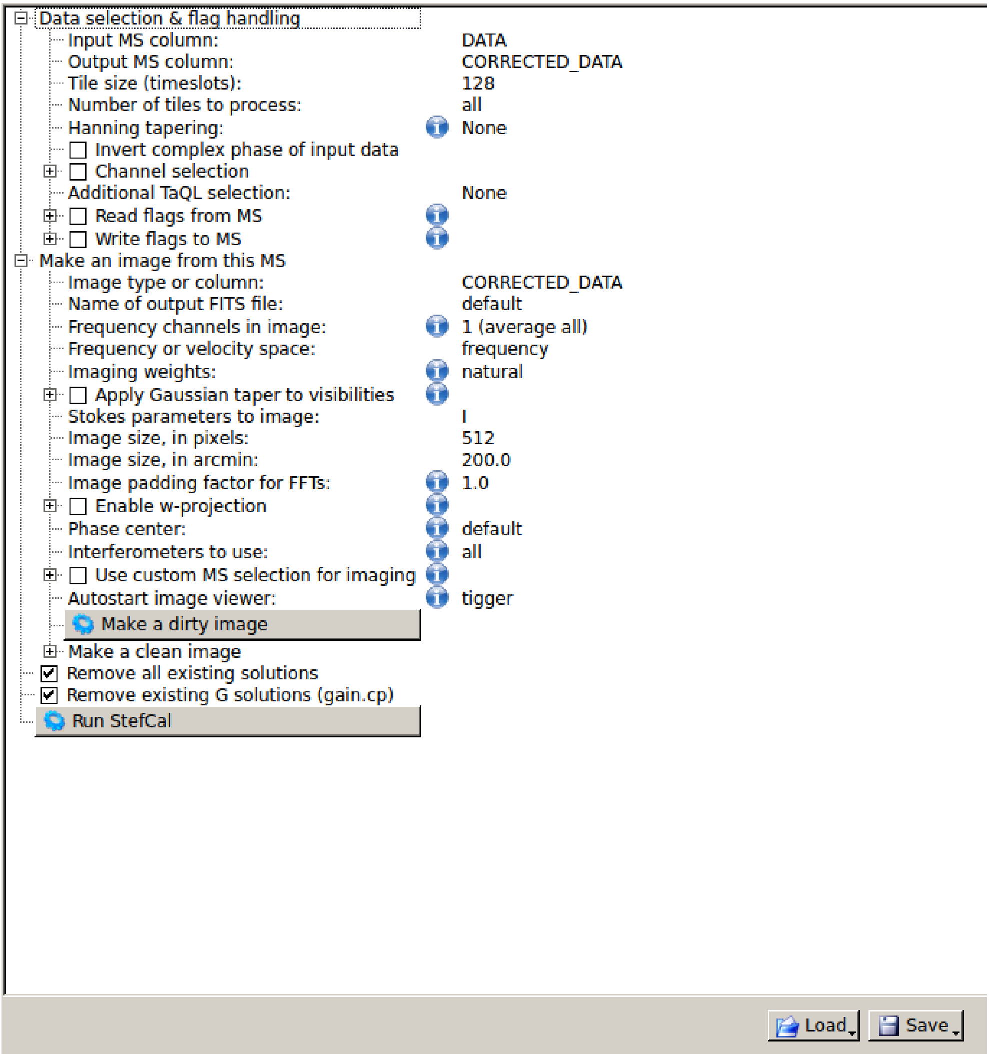

Figure 2: TDL options for imaging purpose. Clicking on Run Stefcal at the left bottom side will give us the input and output visibilities if we open the bookmarks. We need to save the options at this stage so that the tdl-conf.profile will have the correct options for the imaging.

Figure 3: The sources along with the corresponding values.

name	RA	Dec	r	type	I	Q	U	V	RM	spi	shape	tags
PKS1939-6342	19h39m25.03s	-63°42'46.08"	0.0'	pnt	13.2	0	0	0				
M0:A250G	19h37m28.49s	-63°36'39.71"	5.7'	Gau	0.0668	0	0	0			216"x31"@41deg	+cluster_lead cluster=A250 cluster_flux=0.113717 cluster_size=2
M0:A250aG	19h38m11.39s	-63°33'59.12"	1.0'	Gau	0.0469	0	0	0			137"x0"@25deg	cluster=A250 cluster_flux=0.113717 cluster_size=2
M0:B162G	19h39m13.70s	-63°54'20.84"	20.6'	Gau	0.0344	0	0	0			134"x0"@130deg	+cluster_lead cluster=B162 cluster_flux=0.034432 cluster_size=1
M0:C042	19h40m53.58s	-63°13'10.74"	27.8'	pnt	0.0304	0	0	0				+cluster_lead cluster=C042 cluster_flux=0.030417 cluster_size=1
M0:D271	19h35m36.05s	-63°35'32.41"	17.9'	pnt	0.0255	0	0	0				+cluster_lead cluster=D271 cluster_flux=0.025543 cluster_size=1

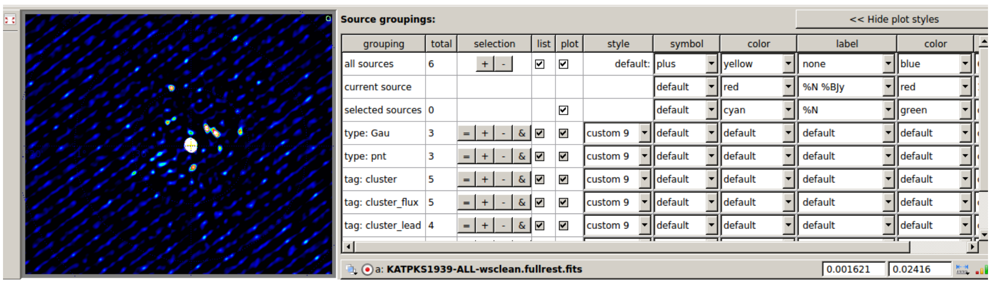

Figure 4: The initial feature of the pipeline for the source extraction.

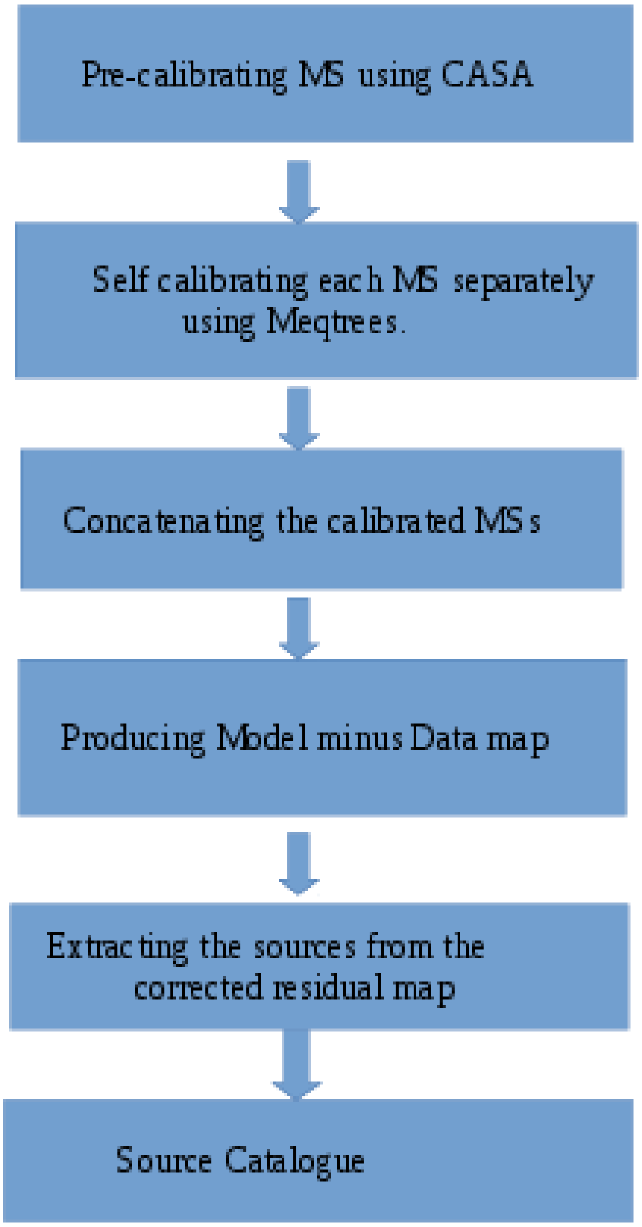

Figure 5: Model Independent feature of the pipeline for the source extraction.

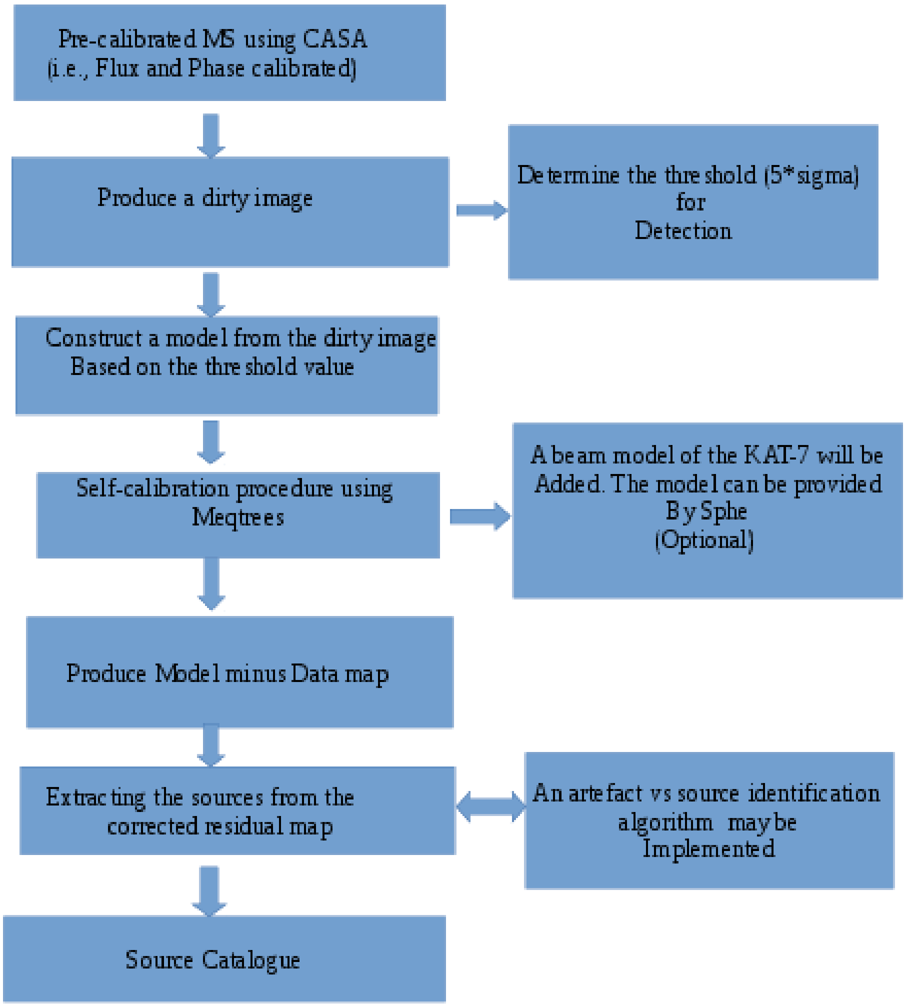

Figure 6: Sources versus threshold for PKSJ1939-6342. Its flux density at 1.83 GHz is ~ 13.249 Jy

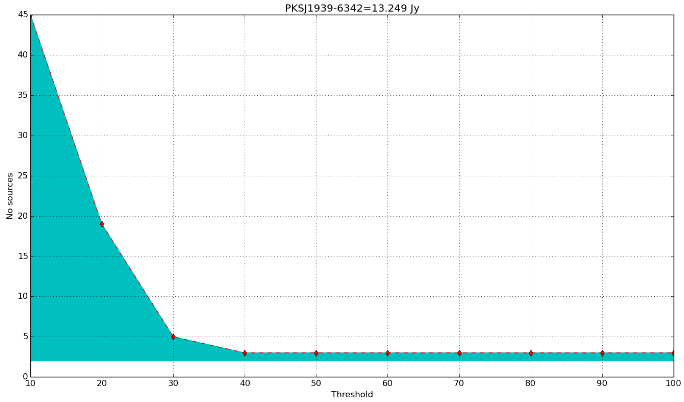

Figure 7: Sources versus threshold for PKSJ0010-4153. Its flux density at 1.83 GHz is ~ 3.64 Jy.

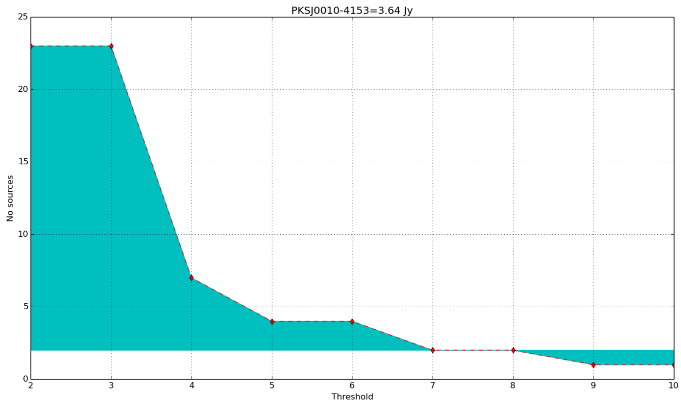

Figure 8: Sources versus threshold for PKSJ0022+0014. Its flux density at 1.83 GHz is ~ 2.43 Jy.