SKA SCIENCE DATA CHALLENGE 1: DATA DESCRIPTION

Overview

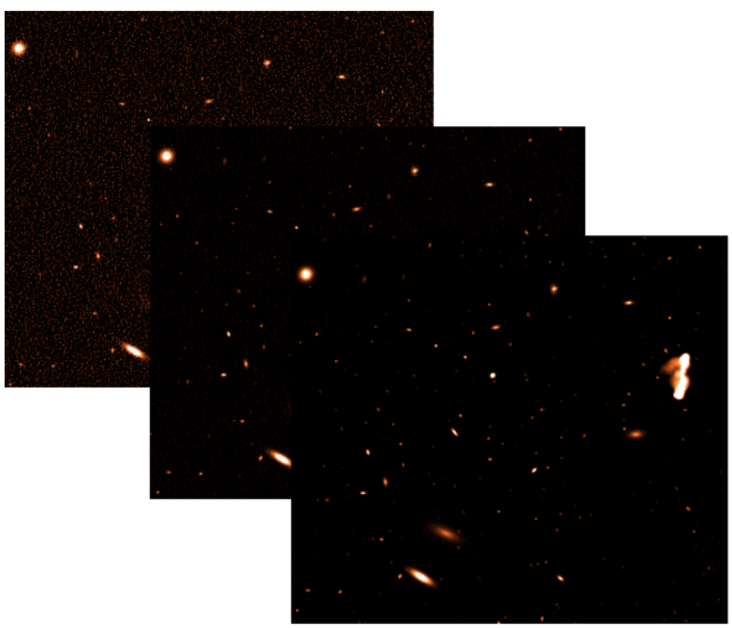
The SKA Science Data Challenge #1 (SDC1) release consists of **9 images** files in FITS format. Each file is a simulated SKA continuum image in total intensity at three frequencies:

1. 560 MHz, representative of SKA1 Mid Band 1
2. 1.4 GHz, representative of SKA Mid Band 2
3. 9.2 GHz, representative of SKA Mid Band 5

Furthermore, three telescope integration depths per frequency are provided:

1. 8 h, representative of a single observation ([[Data Layer]]2/3)
2. 100 h, representative of a medium-depth integration ([[Data Layer]]3)
3. 1000 h, representative of a Deep integration ([[Data layer]]3).

The image below shows the difference between the three different integration hours.



Zoom-in of the 1.4 GHz maps, showing the same region of the sky with different telescope integration: 8, 100, 1000 h left to right.

We notice that the image becomes more accurate as the integration hours increase. Hence, we will be able to find astronomical sources more efficiently. However, this comes with a time cost. All generated images contain the same number of pixel $num Of pixels = 32,768$. However, the field of view (FoV) changes according to the frequency used since the higher the frequency, the smaller the wavelength, resulting in a finer Angular Resolution, therefore:

- 560 MHz --> 0.60 arcsec
- 1.40 GHz --> 0.24 arcsec
- 9.20 GHz --> 0.037 arcsec

And since the number of pixels is the same, this will result in a smaller field of view 'map size' as the frequency increases, resulting in:

- 560 MHz --> 5.5 degrees
- 1.40 GHz --> 2.2 degrees
- 9.20 GHz --> 0.33 degrees

The simulated field is nominally centred at RA=0, Dec=-30 for each map. The sky model is a plausible realisation of the radio sky at those frequencies, but there is no attempt to make it similar to the actual sky at those coordinates. The nine maps share the same sky model realisations, allowing cross-matching between frequencies and direct comparisons between results for different noise levels.

Ancillary Data also accompany the dataset:

- The Primary Beam and Synthesized Beam for each frequency band are provided as FITS format images.

And a training set:

- A “Truth Catalogue” for each frequency band, listing all embedded sources and their properties within 5% of the FoV area, as an ASCII format table. This Catalogue also serves as a submission template for the Data Challenge.

Questions:

- What are the relationships between frequency, wavelength, and wave speed?
- Why are 560MHz, 1.4GHz, and 9.2GHz significant in the SKA?
    - Ask the nearest Postdoc you see

DC1-SKY Model Description

As we mentioned above the sky model is a practical realisation of the radio sky at the following frequencies 560MHz, 1.40GHz and 9.20GHz.
Creating the dataset can be summaries into three steps:

1. Creating the catalogue (just a table for the created sources)

The simulated sky was first generated by [The Tiered Radio Extragalactic Continuum Simulation] (T-RECS). The model generated a catalogue of star-forming galaxies (SFGs) and active galactic nuclei (AGNs), where each source has the following characteristics:

- Integrated Flux density: the total brightness of a source that passes through a surface.
- Sky coordinates
- The Size of the source
- The shape of the source

2. From catalogues to maps

The sources from that catalogue were pushed into a simulated field either as:
- An extended source, if its major axis `size` (defined below) is larger than 3 pixels;
- A compact source if its `size` is smaller than 3 pixels.

The definition of `Size` differs according to the type of the source:

- AGNs: Defined as the largest angular Size (LAS) of the source
- SFGs: The size is defined as the exponential scale-length of the disk.

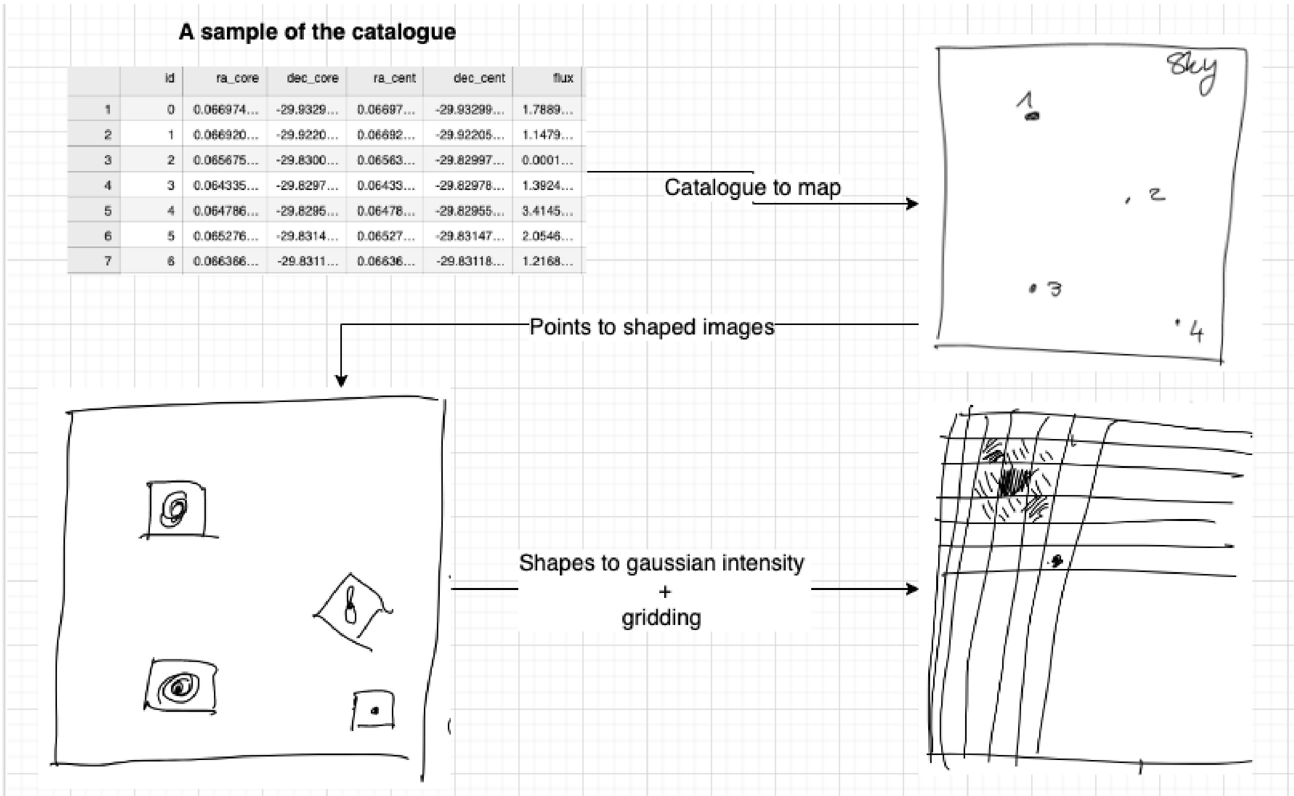
So in this step, there is a focus on the size of the points on the images.

3. Convert point sources to their perspective image shape

In this step, we need to generate shapes for the sources

- For SFGs sources, their shape was modelled after using the exponential Sersic profile depending on their characteristics from the catalogue.
- As for AGNs:
  - Steep spectrum AGNs: A shape was selected from an AGNs library with high-resolution images using the source info. The source was then scaled to its corresponding size and intensity, and then it was randomly rotated.
    - Was the shape randomly selected or using the source's info (from the catalogue)??
  - - Flat-spectrum AGNs were added as a pair of components, where the two components differ in angular size and flux.
    - ORIGINAL "All flat-spectrum AGNs were added as a pair of components; one of milli-arcsecond scale with a given “core fraction” of the total flux density and another of a specified larger size"- no idea how that looks like

Notice: As we mentioned previously, pixel size depends on the frequency. Therefore the same source can be extended or compact depending on the frequency.



	id	ra_core	dec_core	ra_cent	dec_cent	flux
1	0	0.066974...	-29.9329...	0.06697...	-29.93299...	1.7889...
2	1	0.066920...	-29.9220...	0.06692...	-29.92205...	1.1479...
3	2	0.065675...	-29.8300...	0.06563...	-29.82997...	0.0001...
4	3	0.064335...	-29.8297...	0.06433...	-29.82978...	1.3924...
5	4	0.064786...	-29.8295...	0.06478...	-29.82955...	3.4145...
6	5	0.065276...	-29.8314...	0.06527...	-29.83147...	2.0546...
7	6	0.066366...	-29.8311...	0.06636...	-29.83118...	1.2168...

Challenge Description

The challenge set for the community is to undertake:

- Source finding (RA, Dec) to locate the centroids and/or core positions,
  - A popular method for source finding is PyBDSF
- Source property characterization (integrated flux density, possible core fraction, major and minor axis size, major axis position angle)
- Source classification (one of SFG, AGN-steep, AGN-flat)
  - Usually done using machine learning