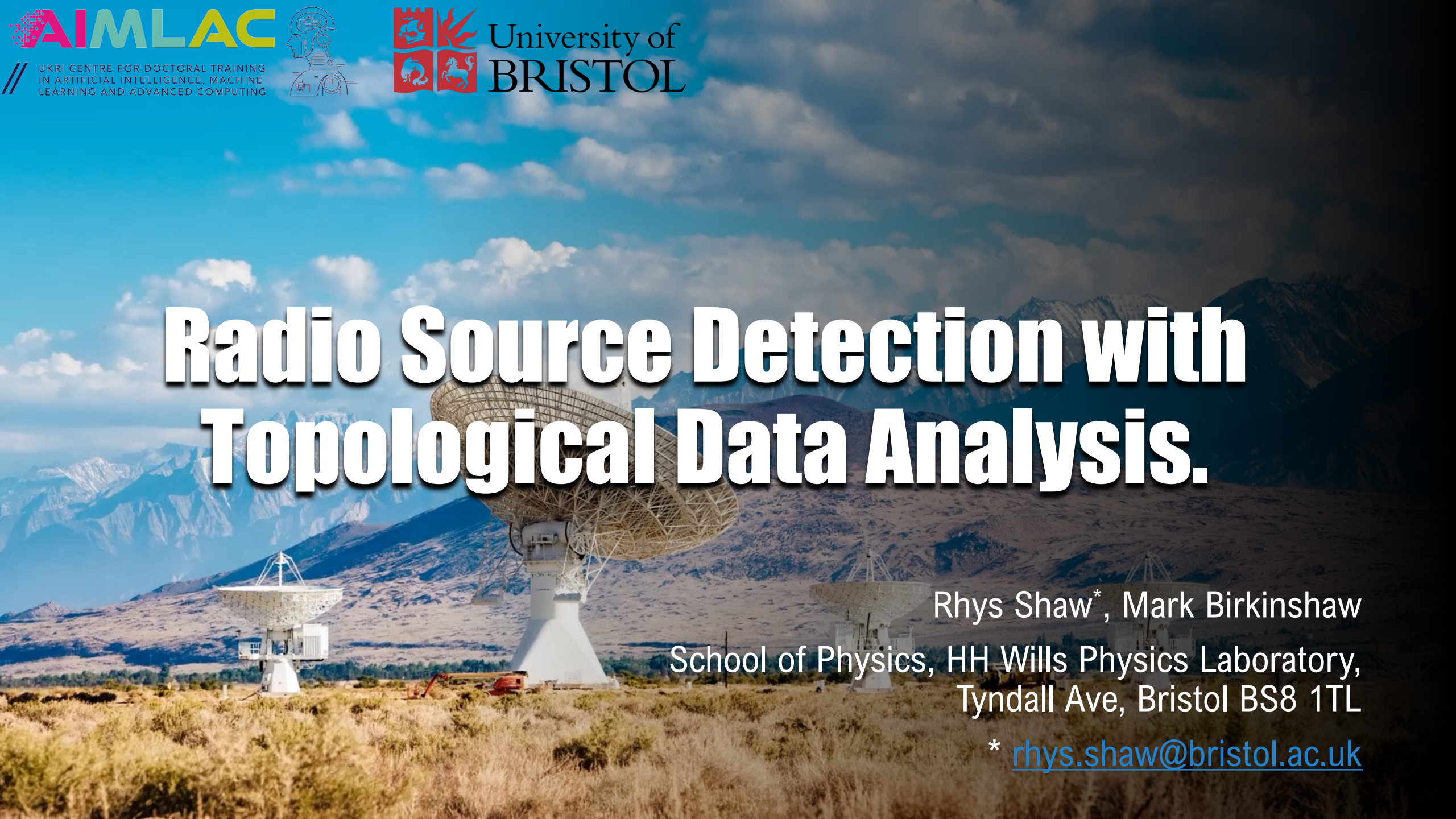


AIMLAC
UKRI CENTRE FOR DOCTORAL TRAINING IN ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND ADVANCED COMPUTING

University of BRISTOL

Radio Source Detection with Topological Data Analysis.

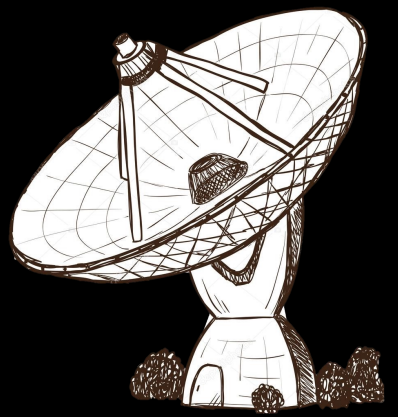
Rhys Shaw*, Mark Birkinshaw

School of Physics, HH Wills Physics Laboratory, Tyndall Ave, Bristol BS8 1TL

* rhys.shaw@bristol.ac.uk

Content

Introduction to Radio Astronomy.
Radio Surveys
The Problem
TRSF
Persistent Homology
First Glance Performance of TRSF.
The Future.

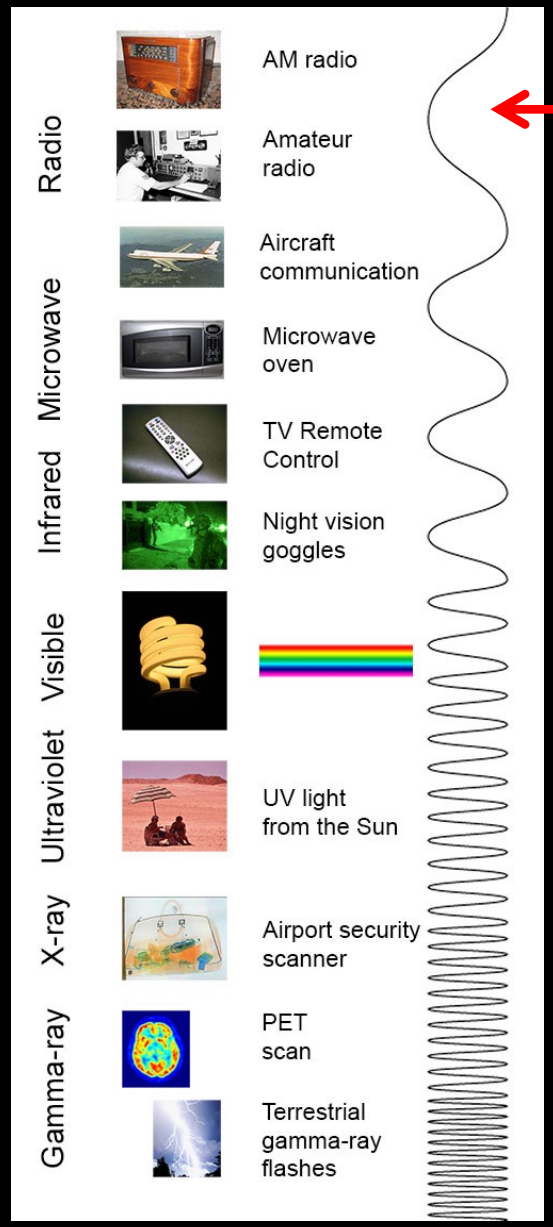


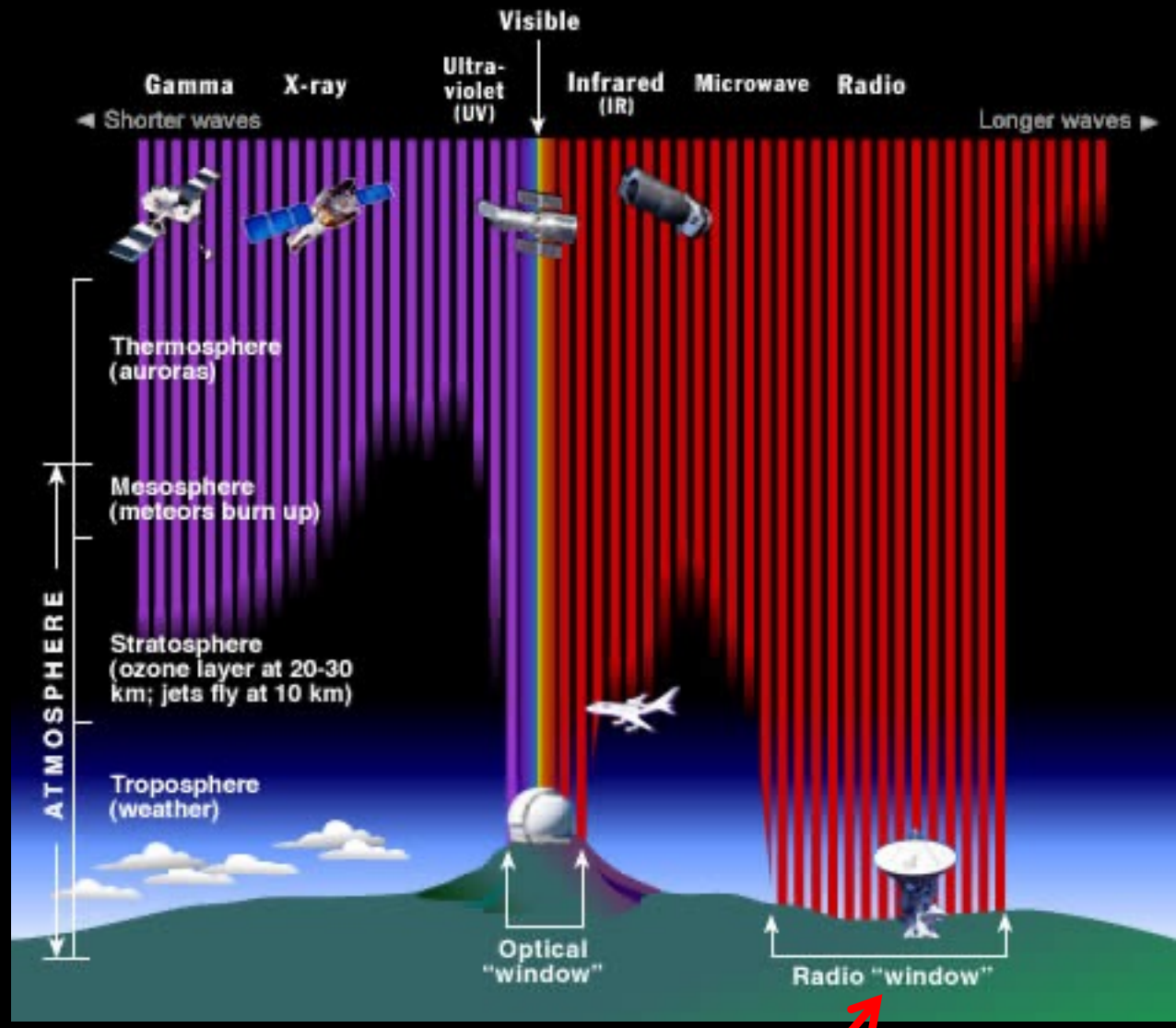
Radio Astronomy

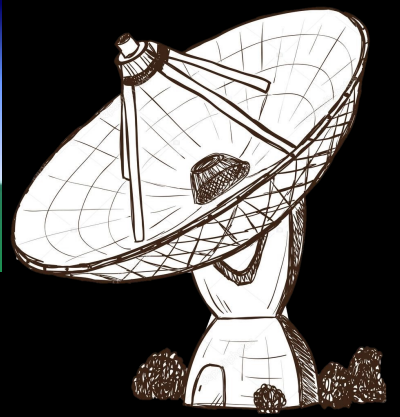
Exactly analogous to optical astronomy. Radio Astronomy is the study of the radio sky and allows for the observation of many galactic and extra-galactic phenomenon.

- Started in the 1930s when Karl Jansky discovered periodicity in radio-telephone static (23hrs and mins – a sidereal day) which pointed to celestial origins.

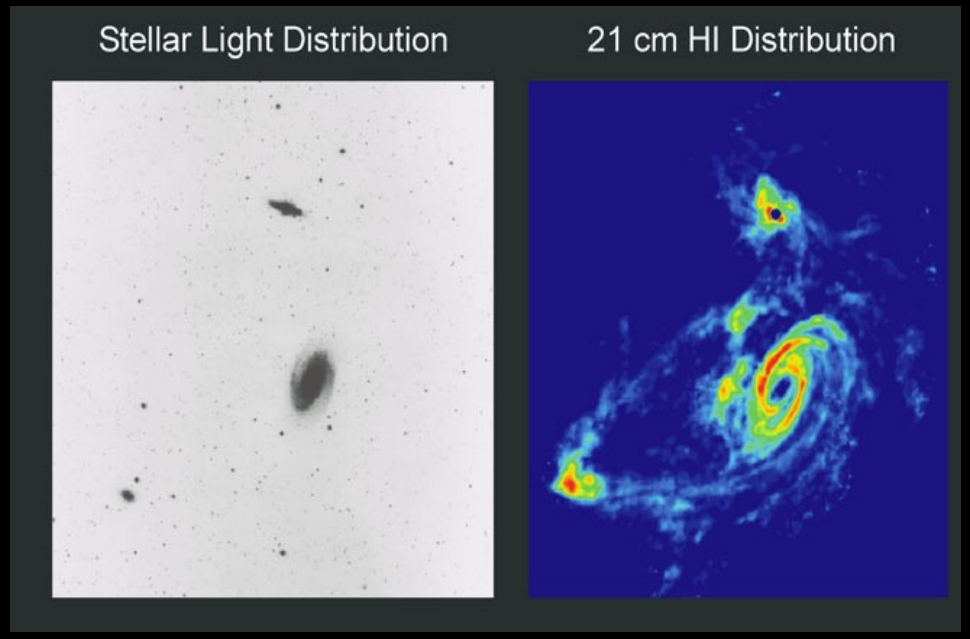
EM Spectrum

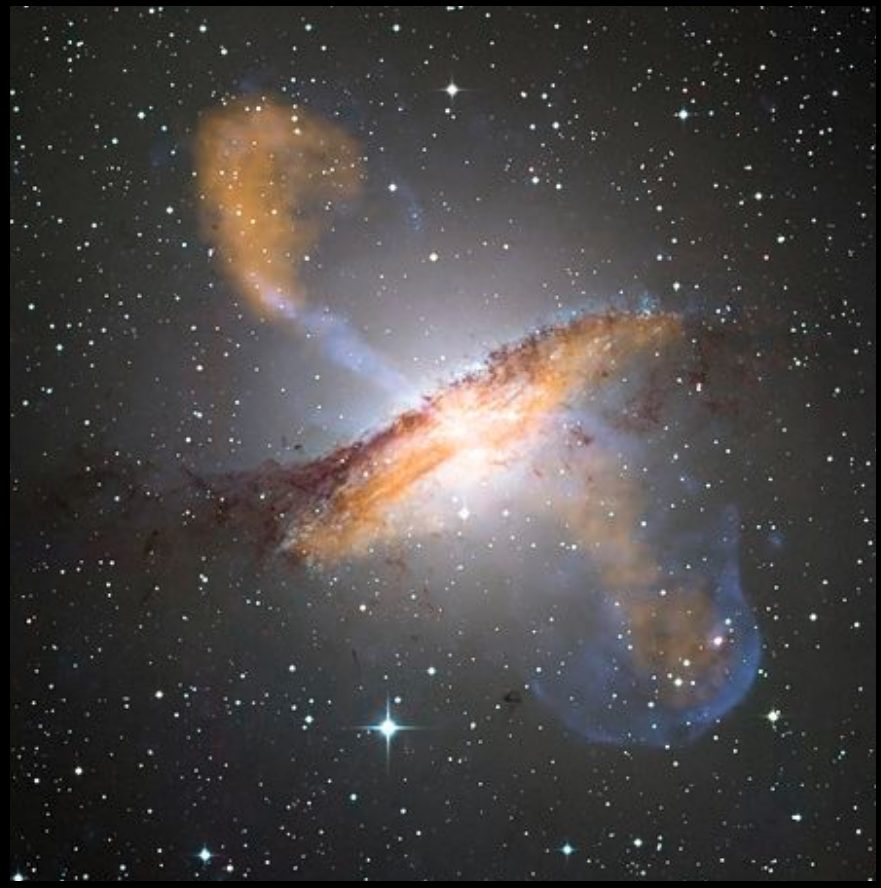

Radio
Microwave
Infrared
Visible
Ultraviolet
X-ray
Gamma-ray
AM radio
Amateur radio
Aircraft communication
Microwave oven
TV Remote Control
Night vision goggles
UV light from the Sun
Airport security scanner
PET scan
Terrestrial gamma-ray flashes


Visible
Gamma
X-ray
Ultra-violet (UV)
Infrared (IR)
Microwave
Radio
Shorter waves
Longer waves
Thermosphere (auroras)
Mesosphere (meteors burn up)
Stratosphere (ozone layer at 20-30 km; jets fly at 10 km)
Troposphere (weather)
ATMOSPHERE
Optical "window"
Radio "window"



Typical Radio Emitters

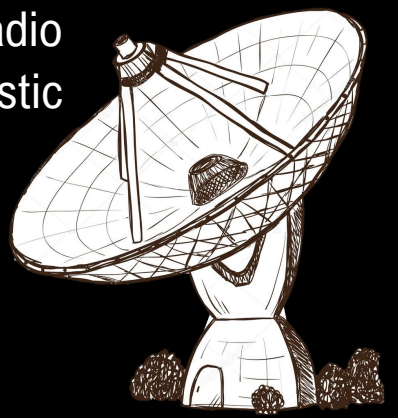




Neutral Hydrogen HI: Special line created by the spin-flip transition of ground state electron (probabilistic unlikely process).

Star Forming Galaxies: Emission comes from synchrotron and free-free emission from HII regions (Ionized hydrogen).

Active Galactic Nuclei: radio emission generated from relativistic synchrotron and free-free emission.

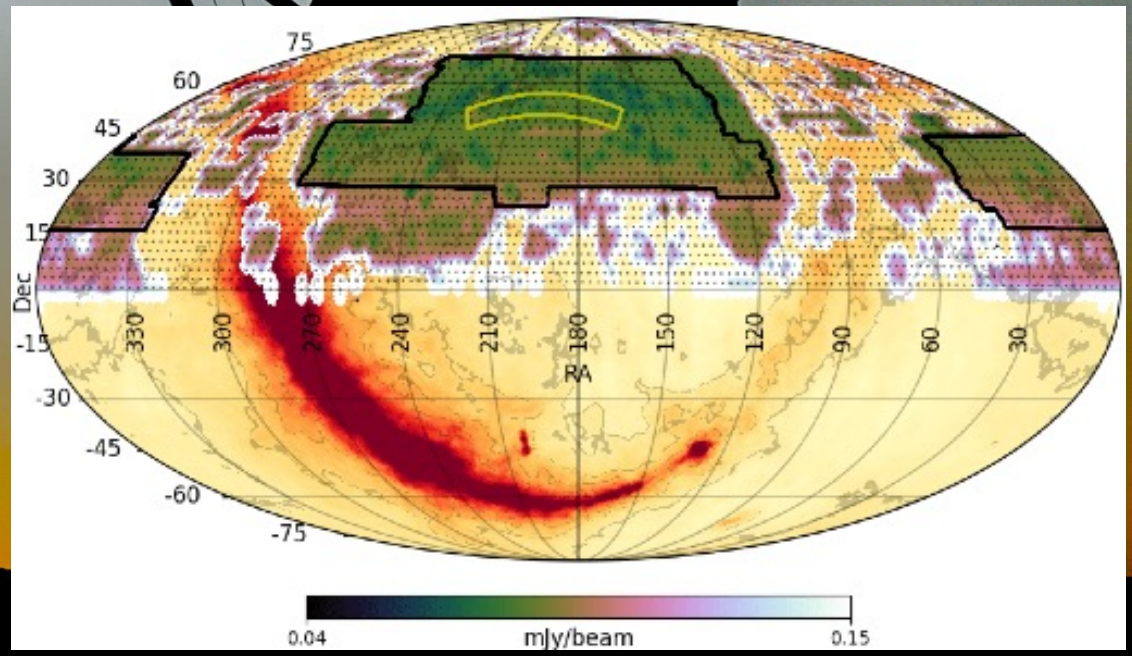


Radio Surveys

Radio surveys are large-scale observations of the radio sky that are used to study a wide variety of celestial objects and phenomena.

LoTSS Survey:
- Survey of the Two-meter sky (120-168 MHz) covering 27% of the northern sky. (Facility is based in the Netherlands)
- > 4,396,228 Detected Sources over 841 separate observations.

SKA (Square Kilometre Array):
- Deep Observations are expected to contain $\approx 10^6$ detectable sources.



Whole radio sky, black boarder marks the area observed by LoTSS (Shimwell et al. 2022).

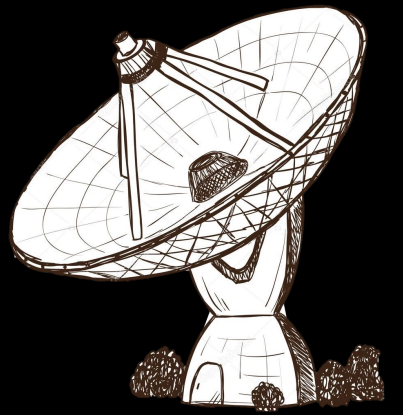
Shimwell T. W., et al., 2022, Astronomy & Astrophysics, 659, A1

The Next Generation of Radio Astronomy

SKA (Square Kilometre Array):

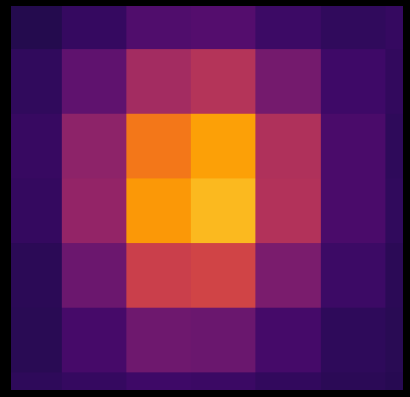
Construction began in late 2022 (Earliest Operational Date 2027)

- Deep Observations are expected to contain $\approx 10^6$ detectable sources.
- This creates a severe data problem; we will be producing more radio maps and catalogues that will be impossible have any significant manual source finding.

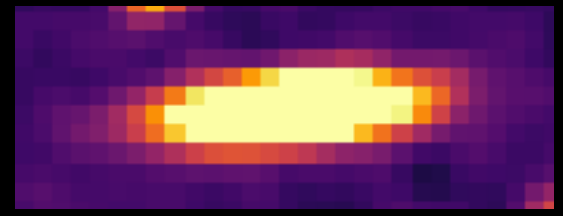


Terminology

Component: - A Single Point-like (gaussian) or extended object. (single peaked)

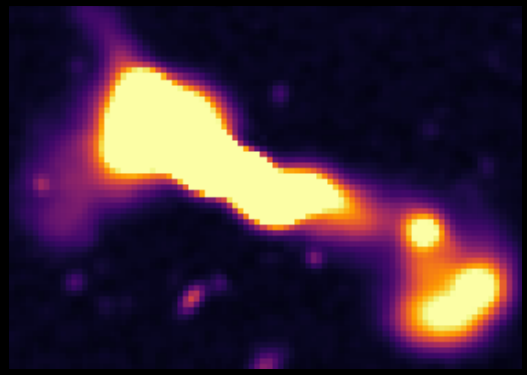


A Gaussian – like point component

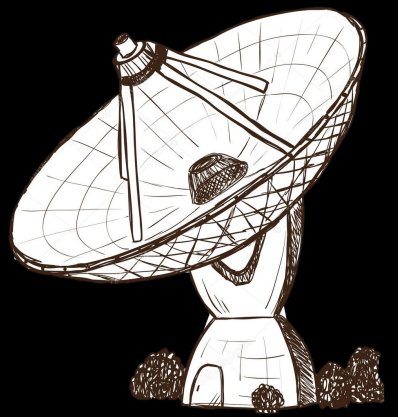


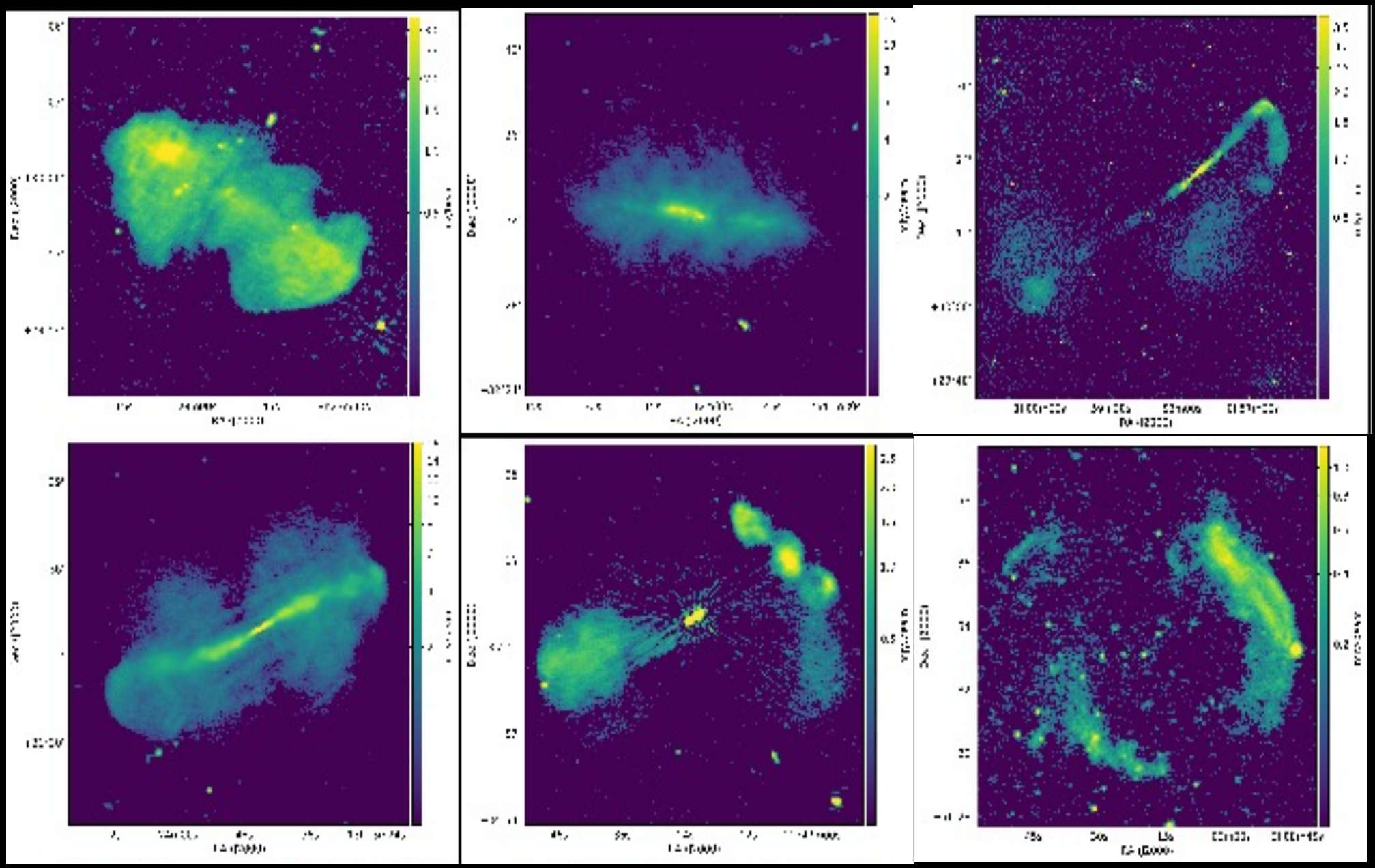
Extended Component.

Source: - A component or group of components that has an astrophysical origin.

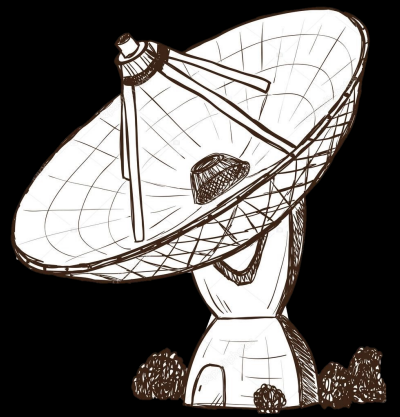


Source with multiple Components.



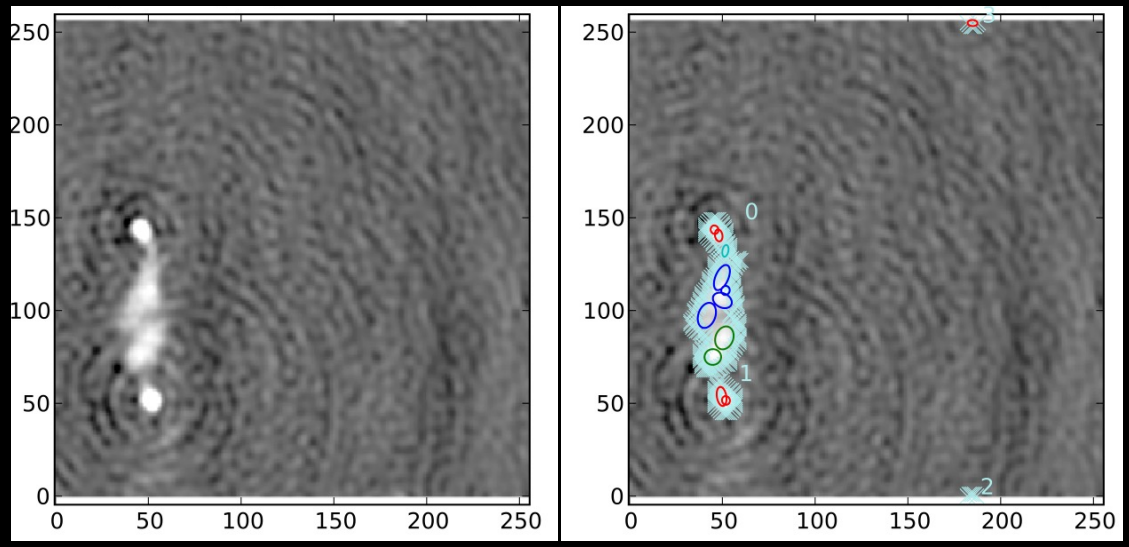


A Collection of Active Galaxies observed in the LoTSS survey (Shimwell et al. 2022)

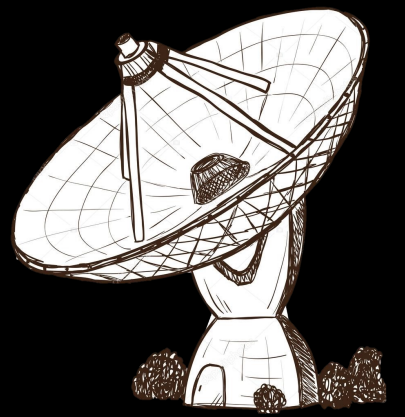


The Problem

- Grouping components of radio sources is important for finding corresponding host galaxies. Traditional methods (typically human inspection) becomes impractical within the new error of radio observatories, that will produce more data than ever before.
- Popular radio source detectors, like **pyBDSF**. Do not preform well on large and source dense images and was not designed for component association



pyBDSF gaussian islands and components for this extended multi-component source.



Mohan N., Rafferty D., 2015, Astrophysics Source Code Library, p. ascl:1502.007

What does a Solution need?

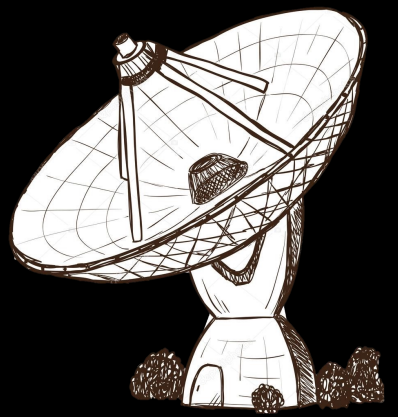
We need:

- A **SOURCE FINDER** that has high completeness of detections and allows for the detection of “all” constituent components (the focus of this talk).

- ASSOCIATION MAKER, Machine Learning Model (Task for the future).

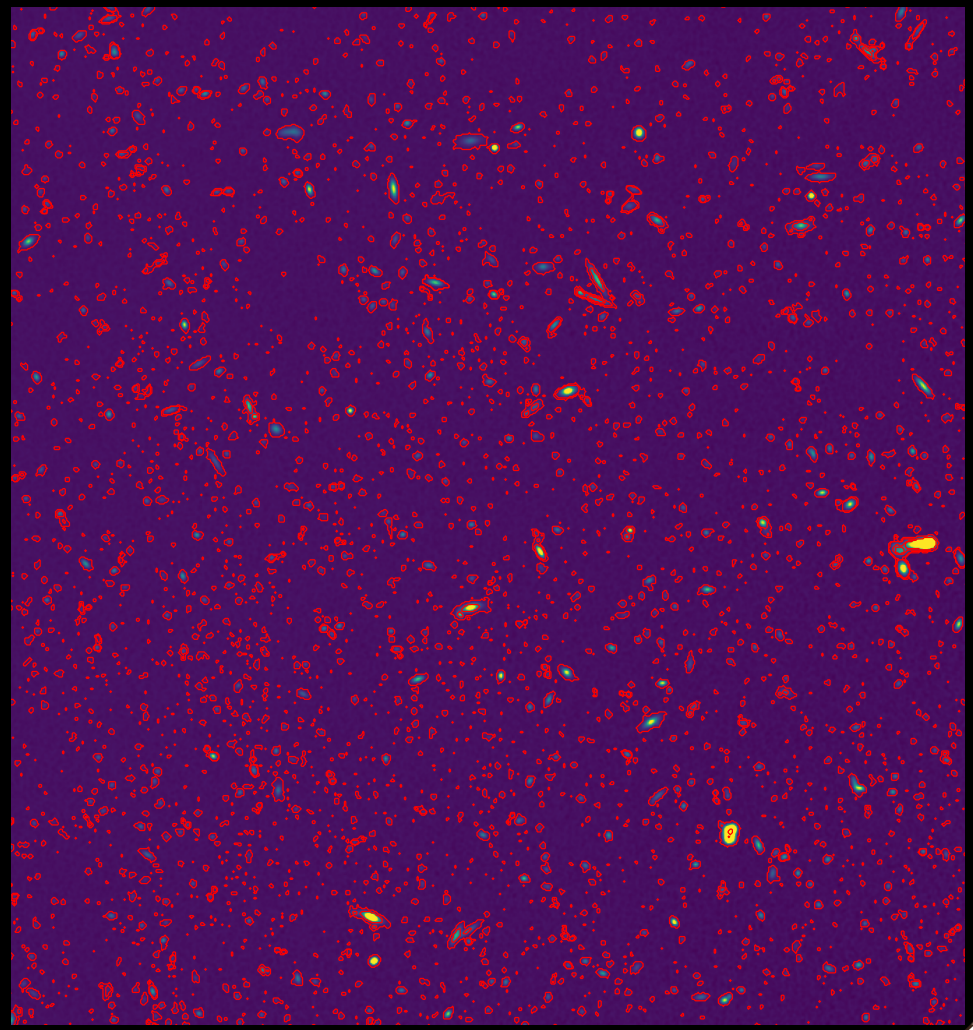
Any Machine Learning solution will require Training Data:

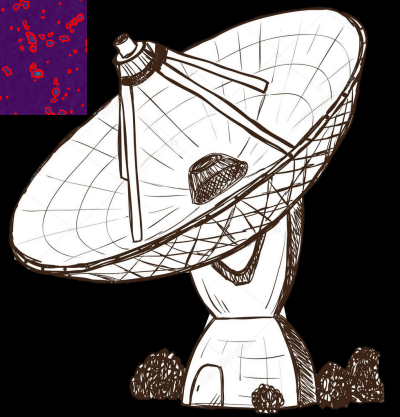
- Simulated Data from the SKA Data Challenge 1.
- Small astronomer created datasets of associations.
- Citizen science association (LoFAR).

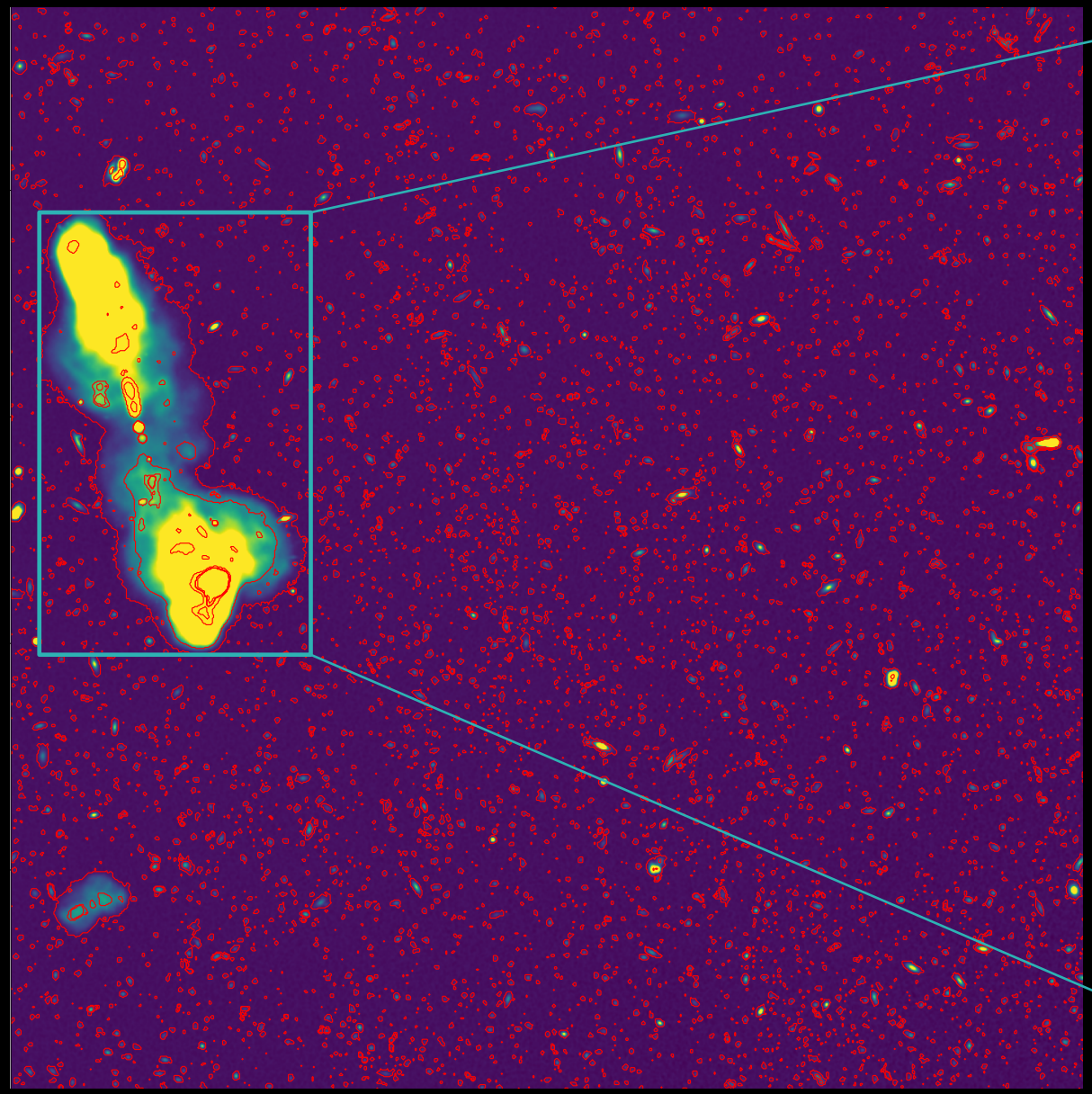


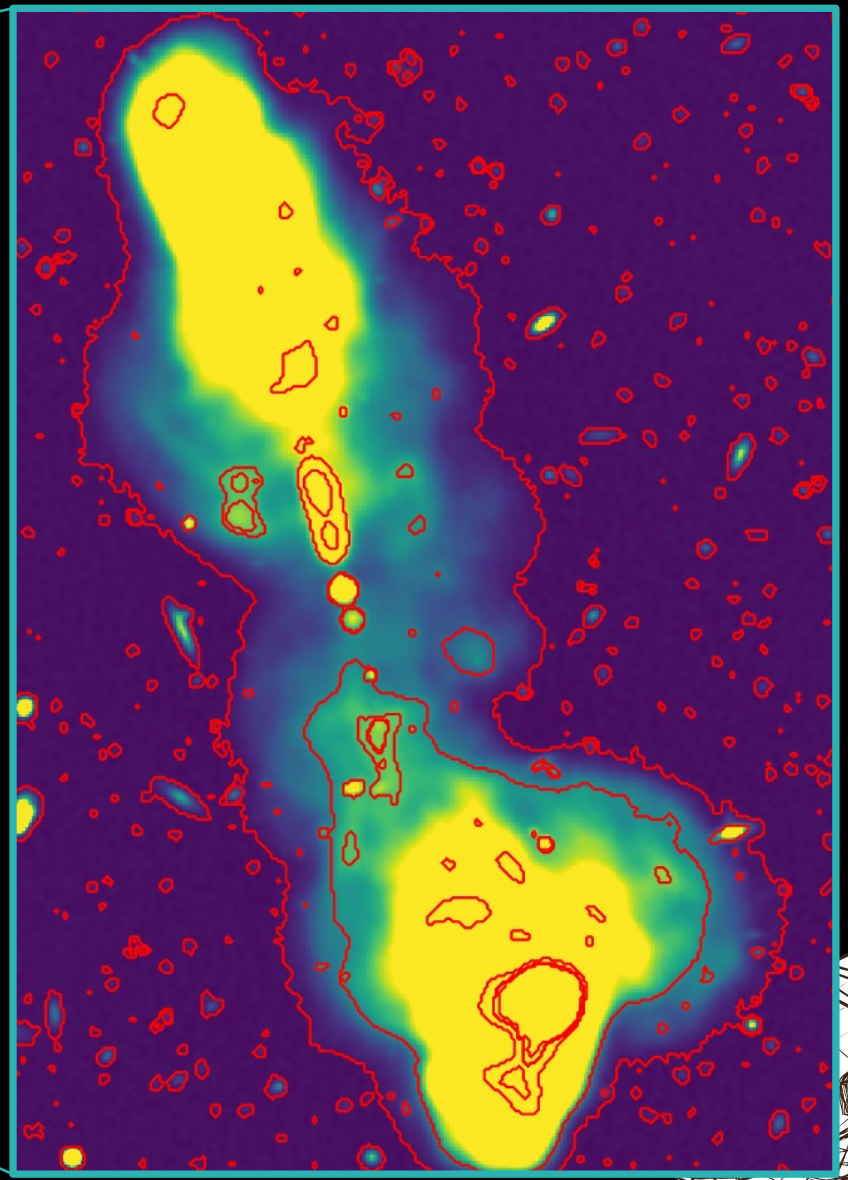
TRSF (Topological Radio Source Finder)

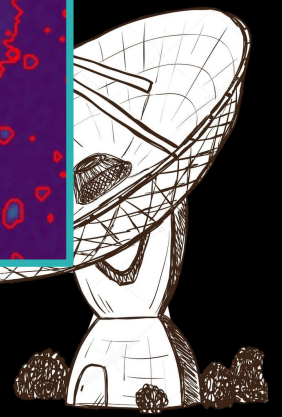
- Generates suggested regions for components using birth and death points from the Persistence diagram (we will see what this is in the next slide).
- Segments sources at all scales even close to the noise.
- Finds and separates overlapping components.











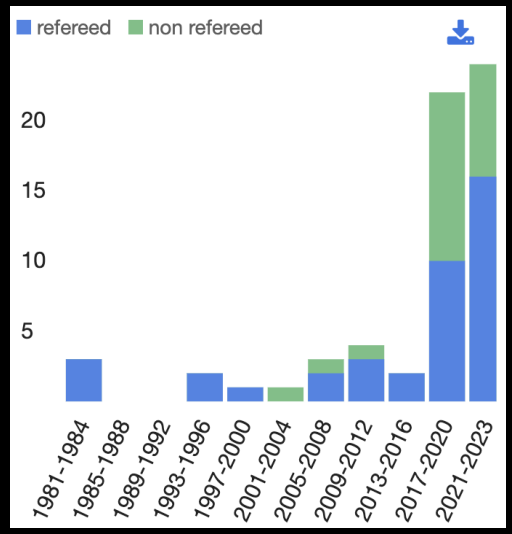
Persistent Homology

A part of Topological Data Analysis

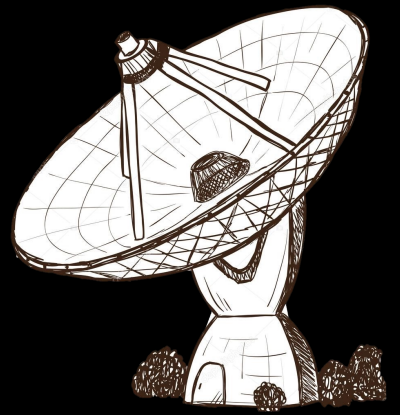
Persistent homology is method used to study the shape and structure of complex data.

It aims to capture essential topological features and identify the persistence of these features across different scales or thresholds.

By analysing how topological features such as connected components, holes, and voids evolve and persist we can learn something from the data.



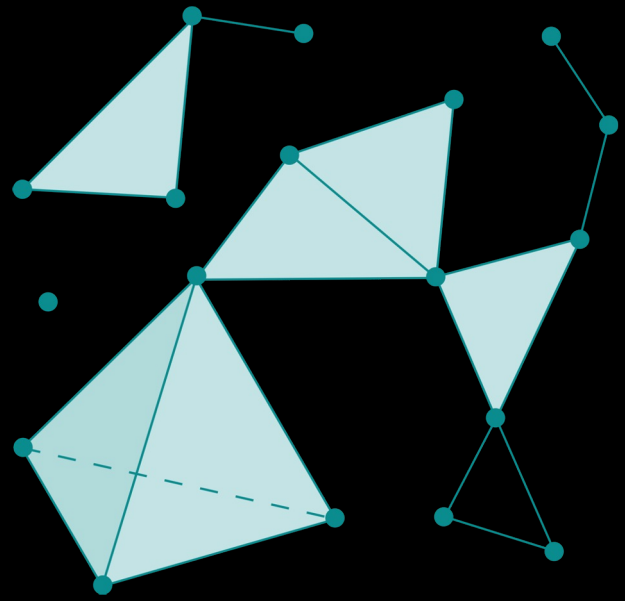
Number of papers added to ADS with Persistent homology keywords.



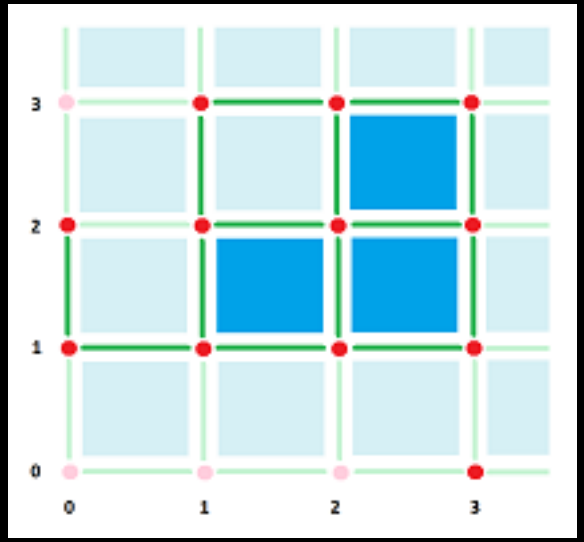
Persistent Homology

To find the persistent diagram of a point cloud it must be represented in a topological space. A simplicial complex is a set composed of points, line segments, triangles and other n-dimensional counterparts.

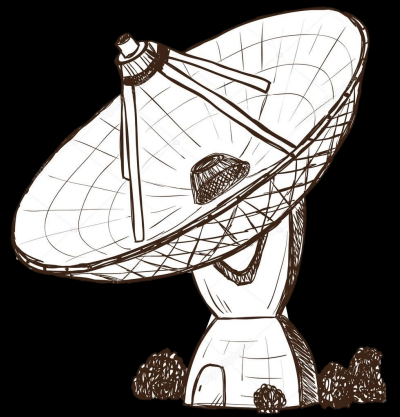
Simplicial Complex



Cubical Complex



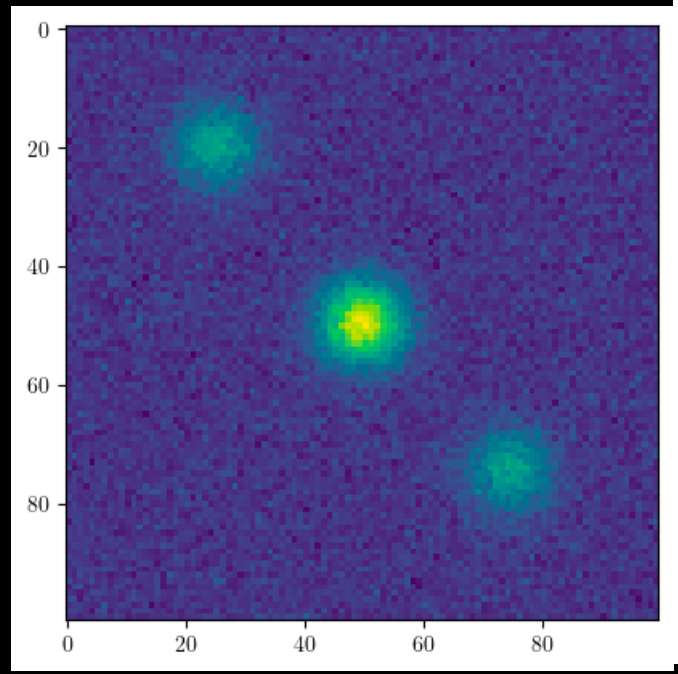
For an Image/2d gridded data we can use a cubical complex. Which is instead a collection of points, line segments, squares, cubes …

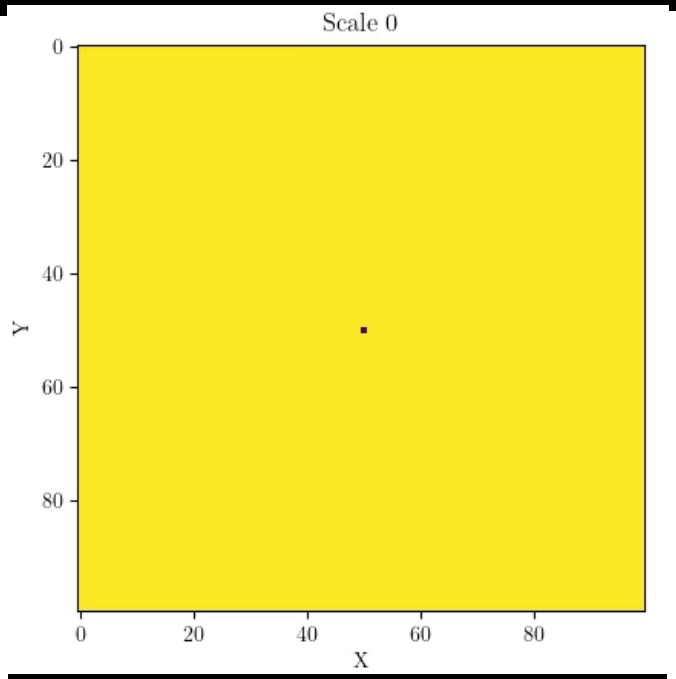


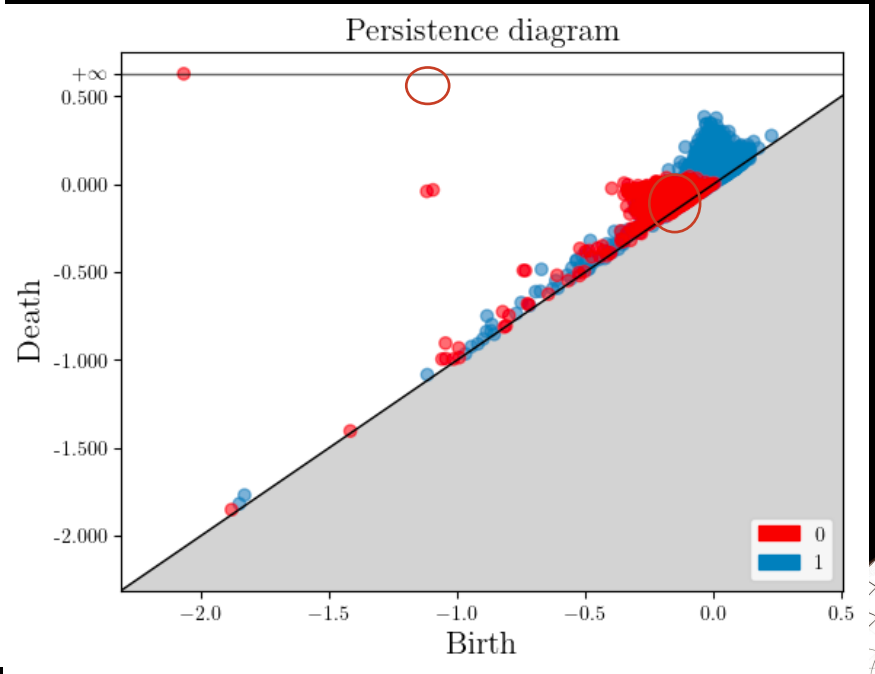
Finding Peaks (0-Dimensional Features)

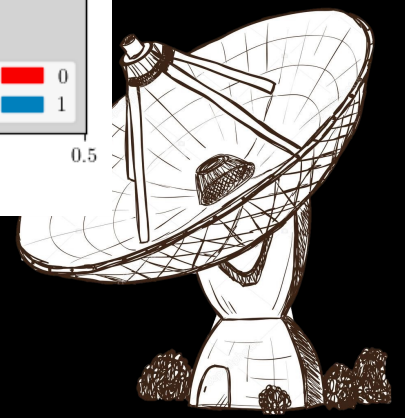
Local minimums are birth points, saddle points are death points in the 0-dimensional persistence diagram. Since we are interested in peaks, we need to invert the image.

Sources



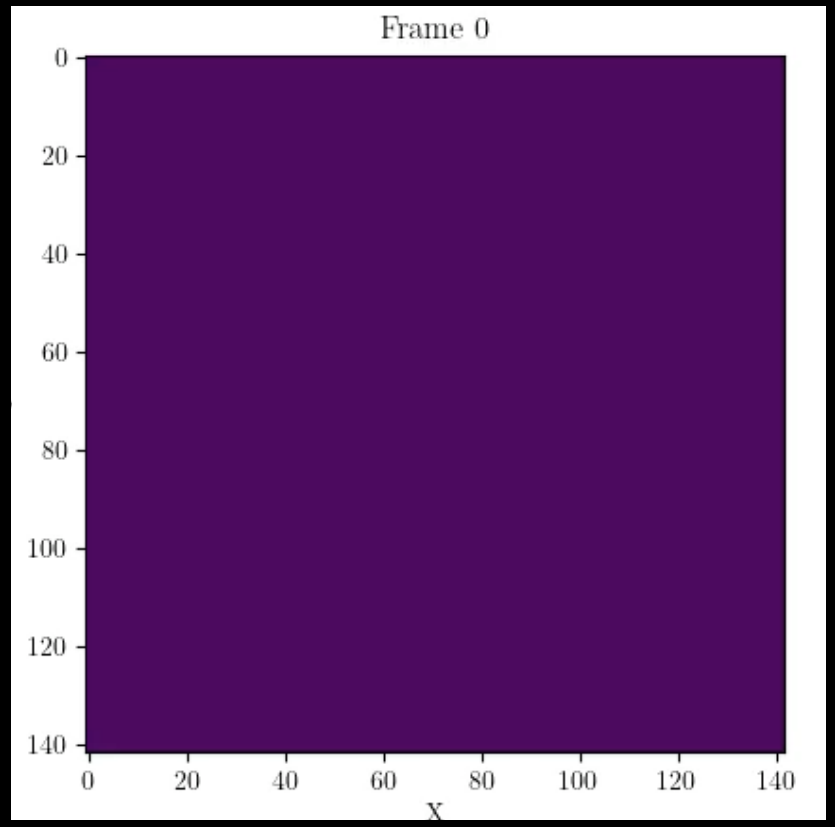


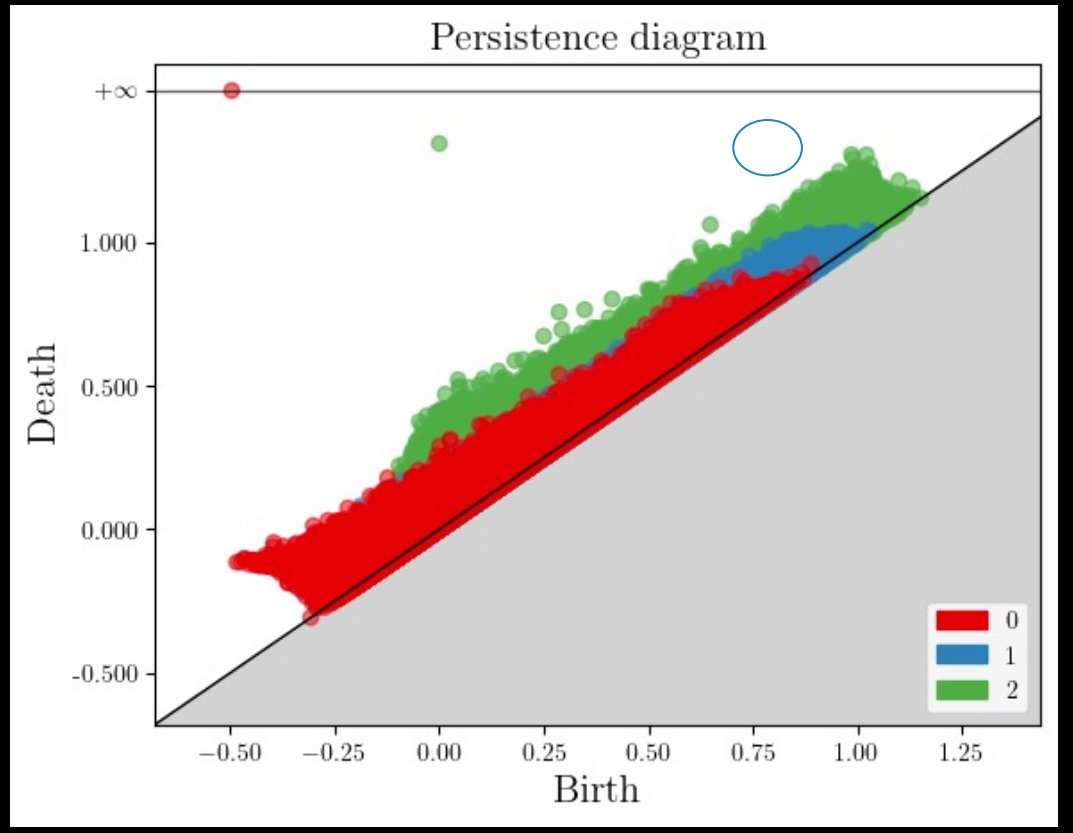


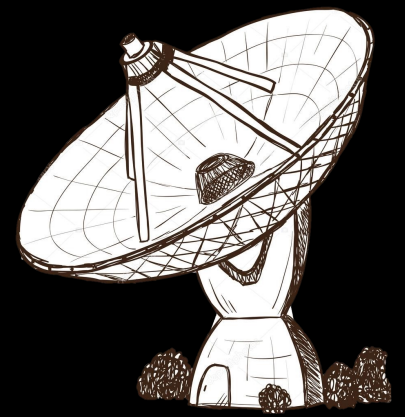


Finding Source in Data Cubes

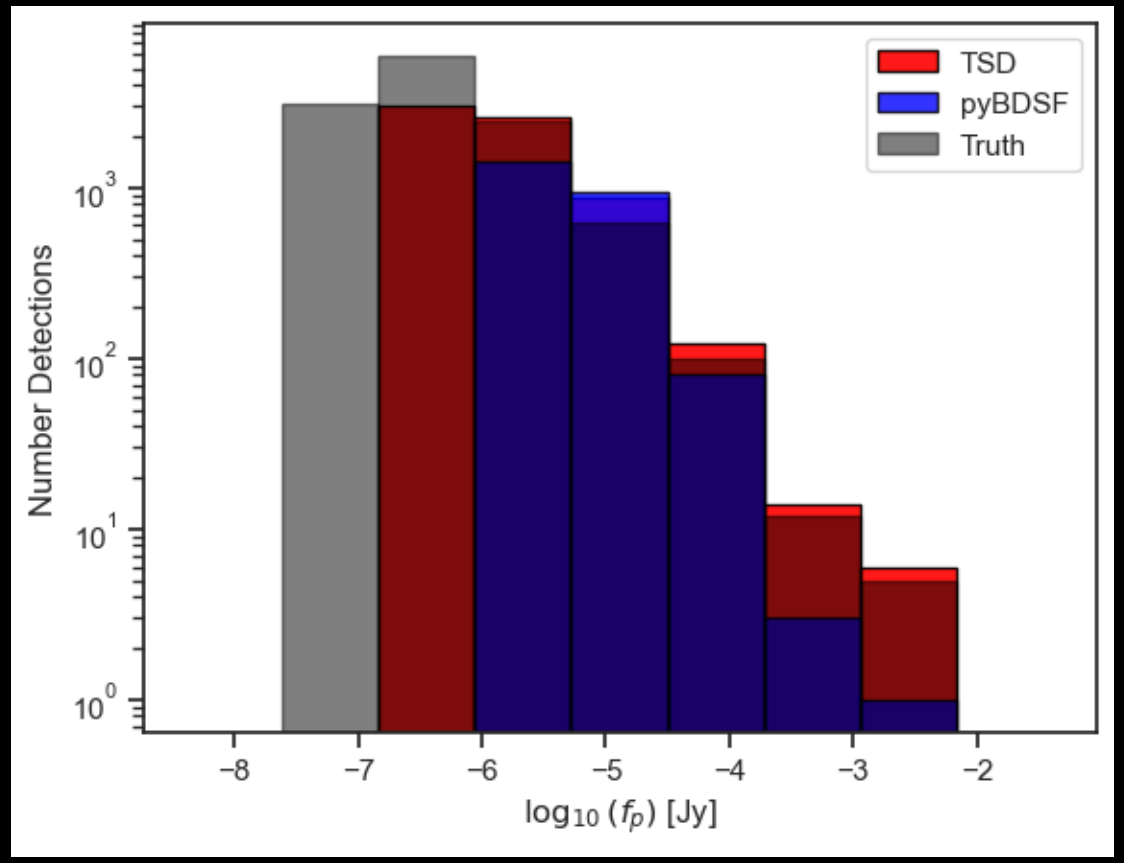
Within a Data cube (2 spatial, 1 spectra dimensions) the 2-dimensional feature would mark the presence of a source that emits in multiple frequencies.



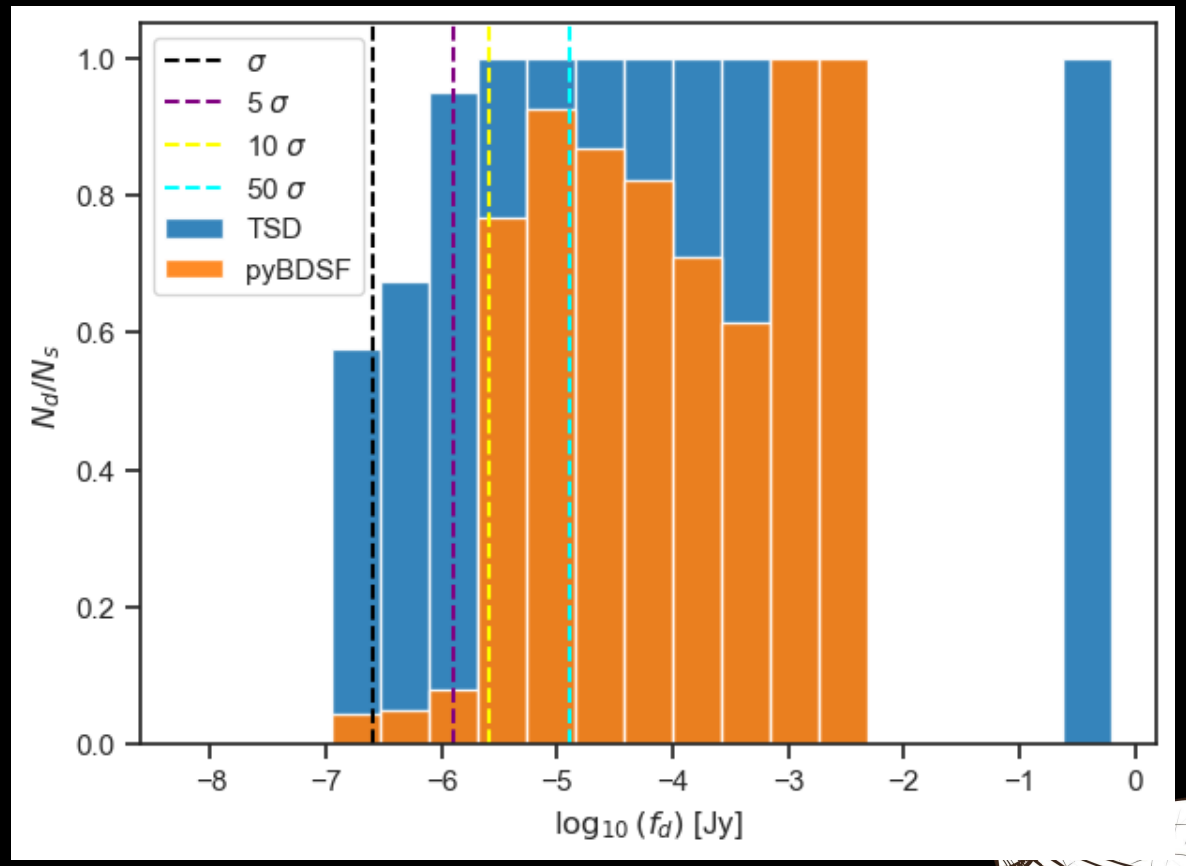




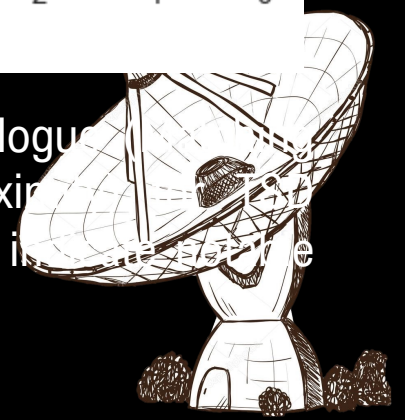
Performance at a Glance



Numbers of detected components within flux density bins, compared to the true distribution (grey) and pyBDSF (blue).



Matched sources compared with truth catalogu made with fitted ellipse overlap and proxi sources and pyBDSF sources. Vertical lines i detection levels.



The Future.

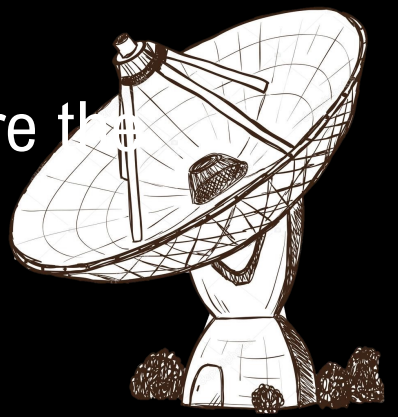
COMPONENT ASSOCIATIONS

- Generate training dataset using components found in TRSF and the truth catalogues in SKA Mid simulated dataset.
- Train a ML model to then classify each component and make groupings of connected components, this should be possible with a mask-RCNN.

- **DATA CUBES**
- Tailoring the TRSF to also work on finding sources within data cubes (in the frequency continuum).

TEST ON REAL DATA!

- Test the source finder on some MeerKAT (an SKA precursor) Data, where the impact of artifacting and other noise sources can be investigated.



Thanks for listening!