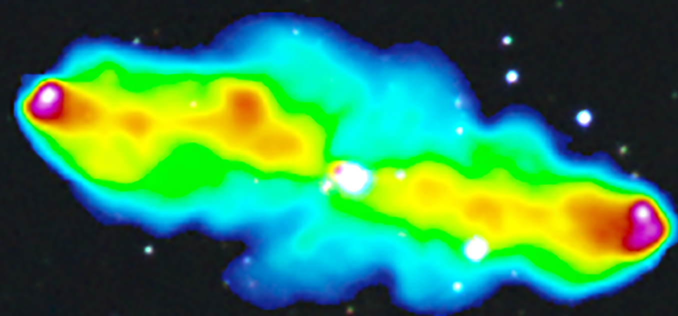
Machine learning and future surveys

Jeremy Harwood



Energetics and life-cycles of radio sources

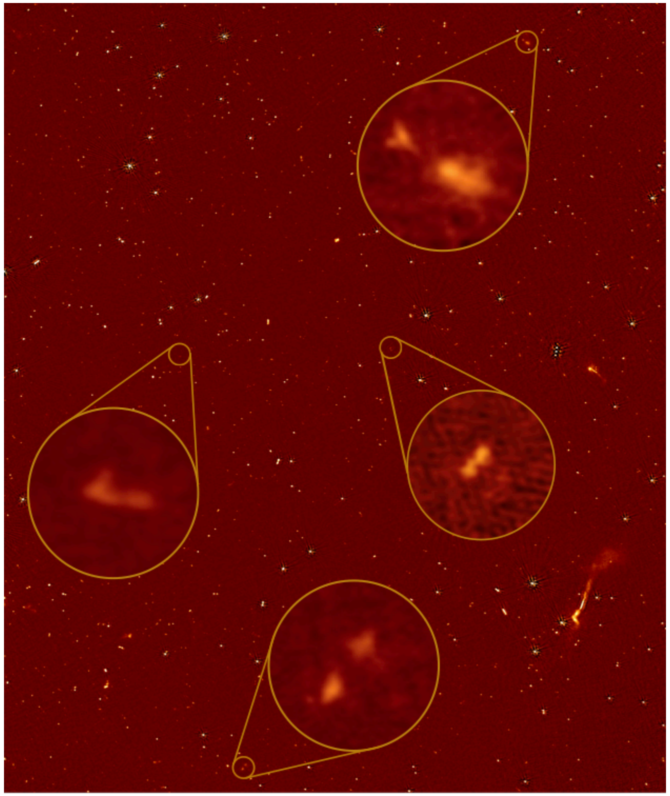
ASTRON, March 2018

Collaborators:

R. Morganti (ASTRON) M. Brienza (ASTRON) Rafael Mostert (Leiden) M. Hardcastle (Hertfordshire)
M. Jarvis (Oxford) A. Stroe (ESO) I. Heywood (CSIRO) L. Morabito (Oxford) K. Duncan (Leiden)

Automated selection to remnant galaxies (and other interesting sources)

- Originally looking to select based on spectral properties e.g. steep spectrum sources
- Many sources aren't as steep as expected at low frequencies (see Marisa's talk for more on this)
- Large surveys are often sparsely sampled i.e. lacking ancillary data for many sources
- May be too high/low to see the spectral break
- Only selecting a small subsample of remnants

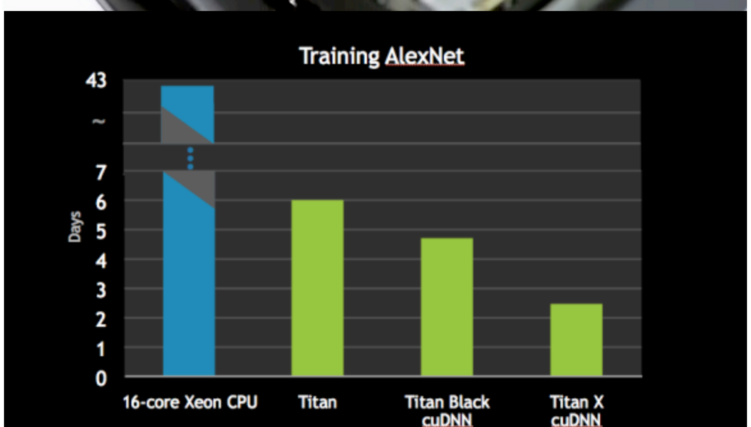


Brienza et al. 2017

How do we find remnants manually?

- Many well studied remnants found serendipitously
- Morphology
  - Relaxed or prototypical FRI/II?
- Compact features
  - Hotspots/inner jets?
  - Radio core?
- Spectrum
  - Steep or shallow?
  - Sharp cut off?
  - Model comparison (CI on vs off)?
- Machine learning is ideal for this!

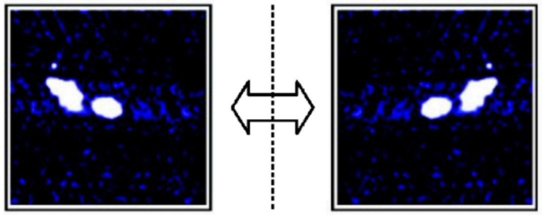
Computer science terminology



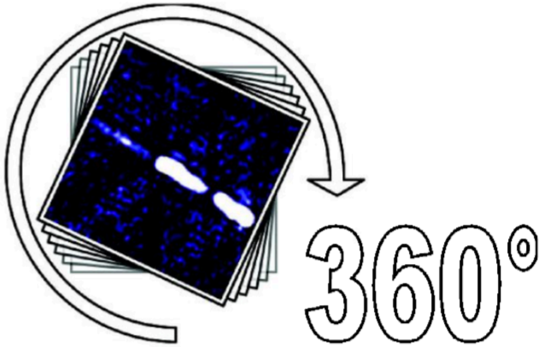
- **Supervised learning**
  - Aim to assign all input sources to a labelled group
  - Used for classification and regression
  - Training set requires labels (e.g. remnant vs. active)
  - Answer questions such as is this an FRI or FRII?
- **Unsupervised learning**
  - Aims to find structure and relationships between groups
  - Primarily for clustering (although many other uses)
  - Can be used to, for example, cluster similar morphologies
- **Feature engineering**
  - What to use and why
  - E.g. total flux, peak flux, or the ratio?
  - Key to making your method a success!

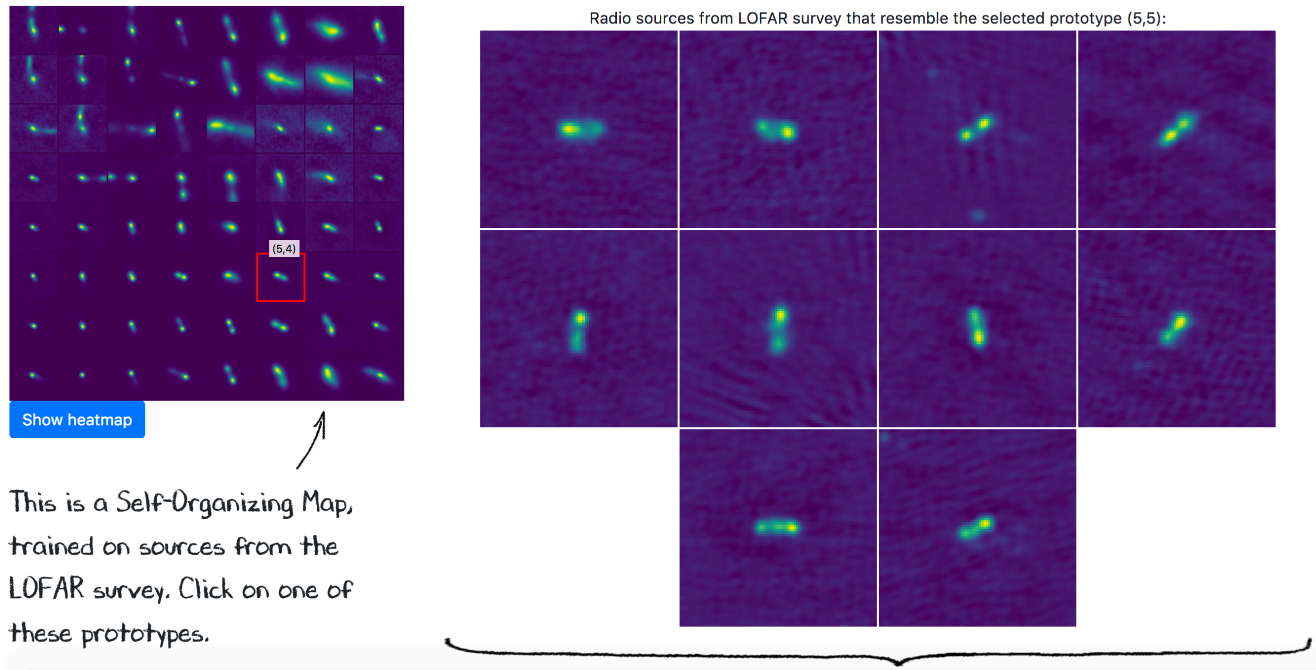
Morphological classification
(Rafael Mostert, Ken Duncan et al.)

- Have used unsupervised learning in the form of an artificial neural network
- Uses PINK: a rotation and flip invariant ML algorithm (Polsterer, Gieseke, and Igel, 2015)
- Computationally expensive; requires GPU computing to be completed in a reasonable time
- Created a Self-Organising Map (SOM) for the Lockman Hole
- Can be mapped to other fields in order to identify similar objects (distance from prototype)



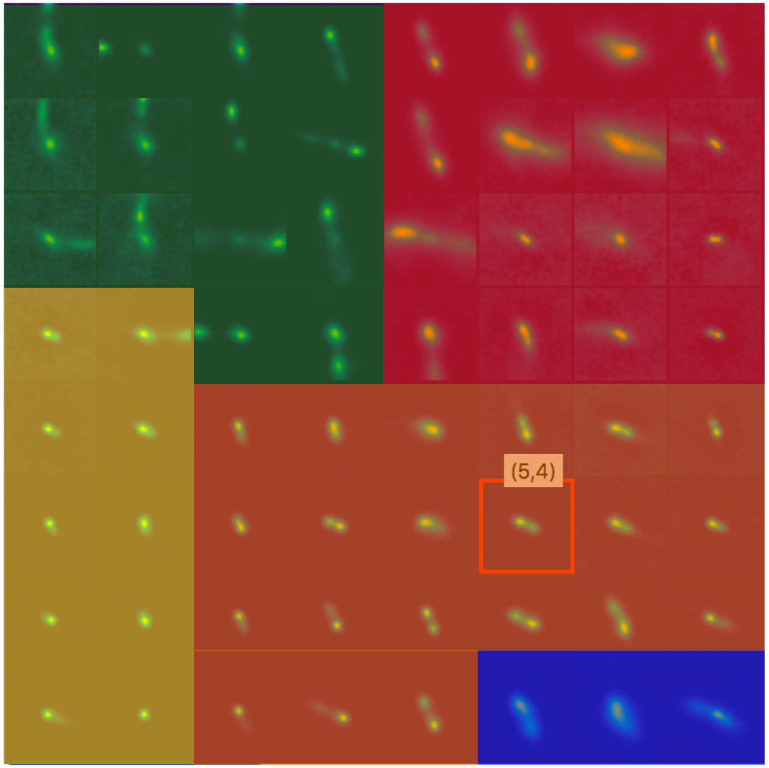
Polsterer, Gieseke, and Igel, 2015





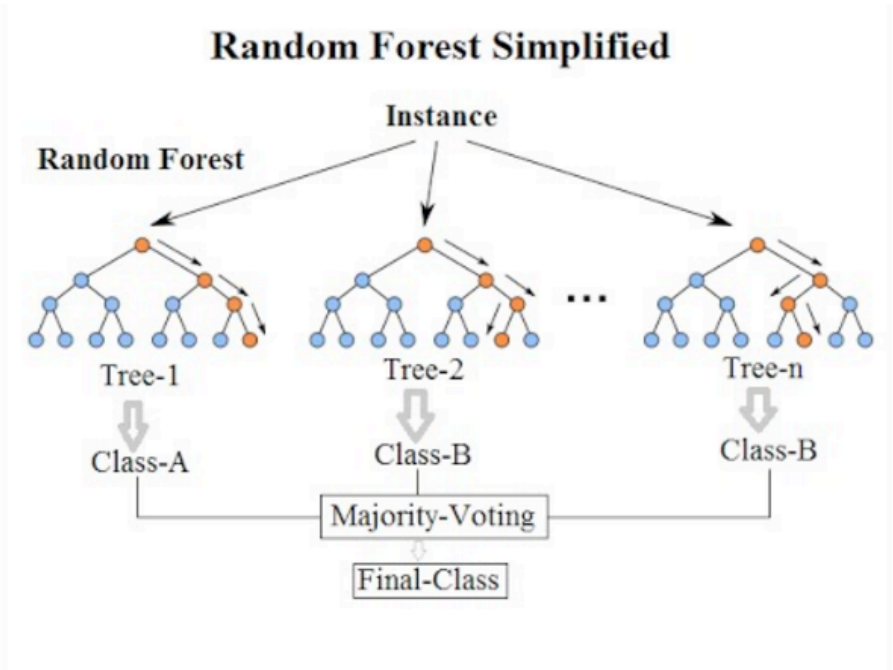
- These prototypical sources can be mapped to other fields (e.g. HETDEX)
- Preliminary tests indicated it might be possible to locate some remnant candidates based on morphology alone
- Likely to only be a specific subset e.g. 3C28 is likely a remnant, but very hard to tell from only its morphology, we need more information!

Clustering and morphological classification



- Clustering is the key to unlocking the power of the self-organised map
- Group the map in to N categories based on their morphology
- Maybe able to classify FRI / IIs but this may be secondary for our purposes
- Compact, extended, relaxed etc. is what we want to know
- Importantly, clustering will provide binary and/or analog parameters for morphology that can feed in to the bigger picture

Random forest algorithm

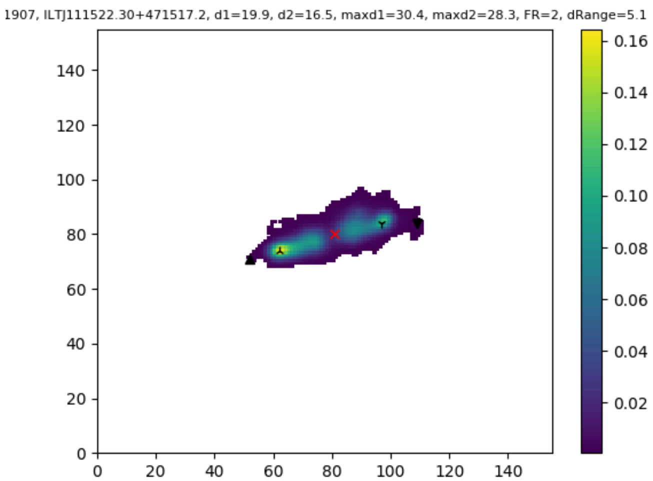


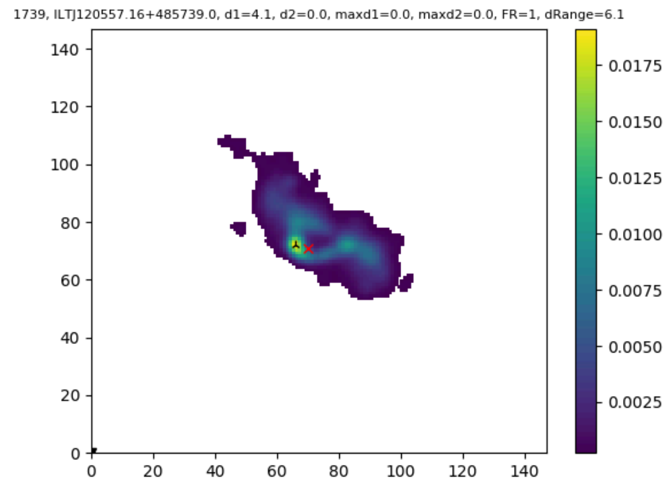
- Need to use all the information we have about sources to unlock the power of ML
- Singular decision trees are fairly common in astronomy
- Simplest (and fastest) solution, but not convinced it is the correct one long term
- Algorithms such as random forest are far more commonly used in the (non-astro) ML world due to reduced over fitting
- This is more complex and requires greater thought wrt to the parameter space

ASTRON Netherlands Institute for Radio Astronomy

LOFAR

Feature engineering for remnant sources

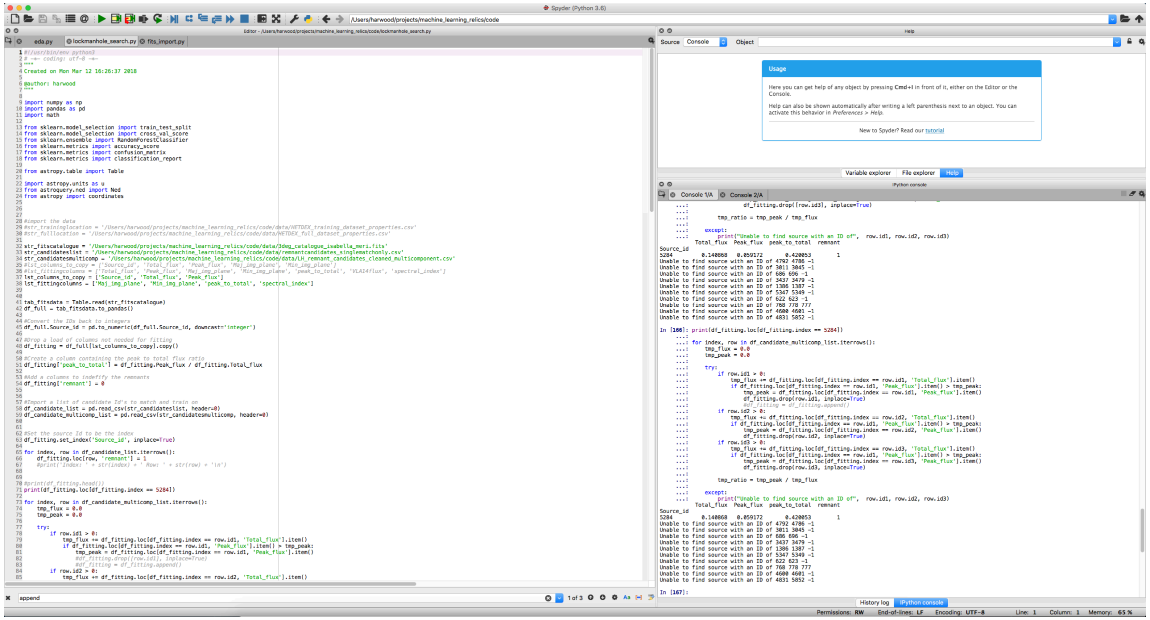




- **What features (observable properties) can/should we use to find remnants?**
- **Known impact: spectral index, CI on/off model fitting, morphology**
- **Unknown impact: peak/total flux ratio, redshift, FRI/II classification, environment**
- **Code such as FRMAL by Beatrix Mingo (pictured above) may be more efficient for FRI/II**

ASTRON
Netherlands Institute for Radio Astronomy

LOFAR



- **Development of the code and feature engineering is underway**
- **Looking at the Lockman hole using the known candidates. This will expand to include a wider variety of remnants as we begin to better understand the parameter space**
- **Main problem: a small training set! Cross verification is currently not possible**
- **One possible solution to this is to combined fields with known candidates e.g. H-ATLAS**

RadioLife

VLA GHz survey of extragalactic sources in the XMM-LSS field (AGES - XL)

- 30 of the brightest extended sources in the XMM-LSS field
- 97 hours split between L- and C-band at ~1 arcsec resolution
- Large amount of ancillary data available (e.g. LOFAR, XMM, VIDEO, DES)
- One of the MeerKAT MIGHTEE fields, and lower resolution (but wider field) survey by Ian Heywood (Oxford)
- Help develop methods for radio galaxies the MIGHTEE surveys

AGES-XL: Science goals and progress

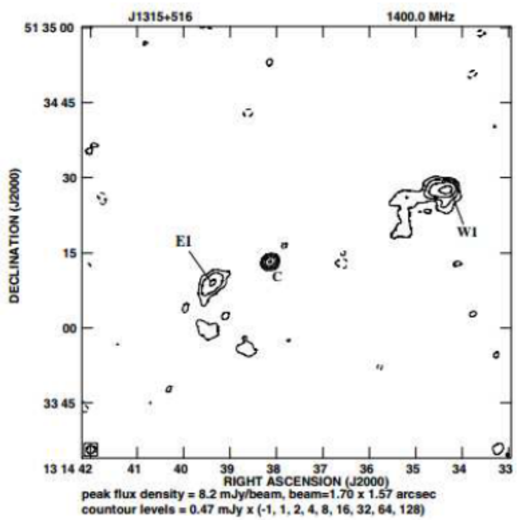
- **Primary science drivers:**
  - What are a robust set of parameters for models of spectral ageing for the radio galaxy population?
  - If the hotspot/lobe discontinuity common in the population?
  - Are recently revised energetics robust?

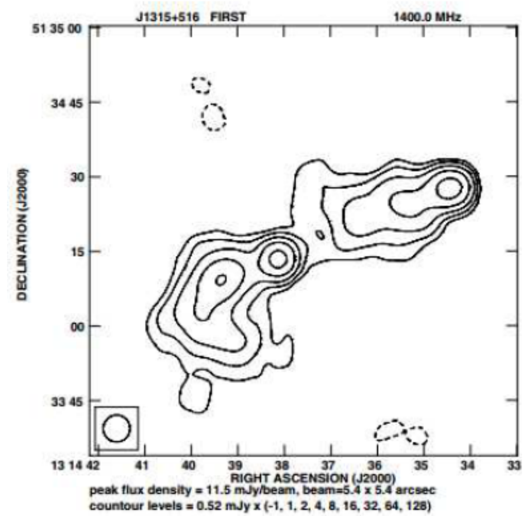
- **In conjunction with other surveys (e.g. MeerKAT, LOFAR) this will help to solve:**
  - What is the characteristic age and lifespan of nearby radio galaxies?
  - How does this age vary as a function of redshift?

- **73/97 hours observed, final observations will be taken summer 2018**

- **Plenty of opportunity to get involved either with the planned science or your own projects!**

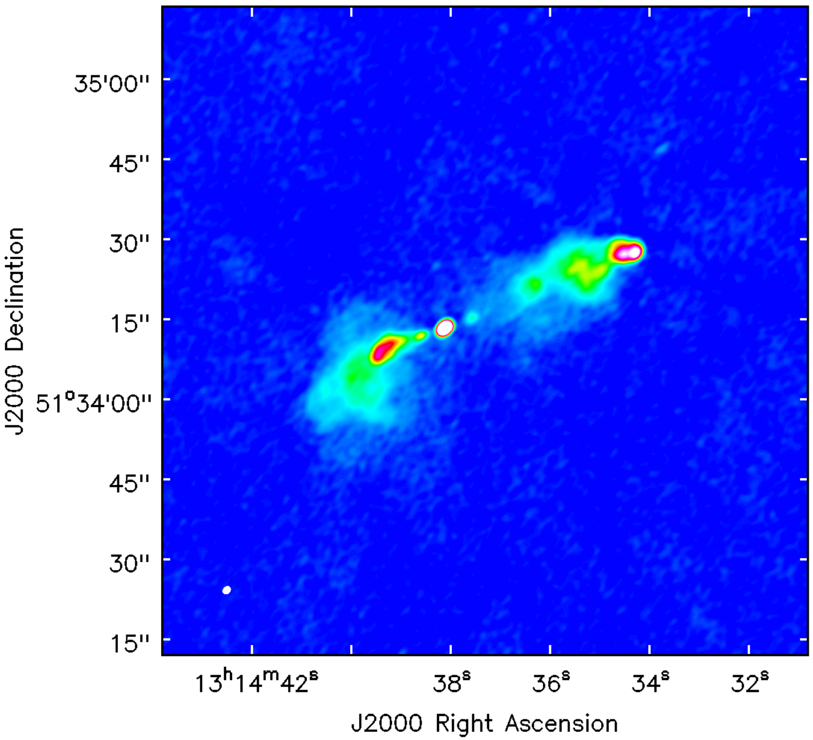
Hybrid Morphology Objects (HYMORS)





- **Discovered by Gopal-Krishna & Wiita (2000) and observed as FR I on one side and FR II on the other**
- **Very few studies have been performed to date. Recently, only a Galaxy Zoo search by Kapinska et al. and a new object found by de Gasperin (2017)**
- **No detailed spectral studies of the original samples or any broadband observations that capture both compact and diffuse emission**

Hybrid Morphology Objects (HYMORS)



JVLA observation of J1315+516 (z = 0.63) at 6 GHz
Full bandwidth between 4 & 8 GHz (C-band)
1.65”x1.25”resolution, combined B&C config.

- **5 examples discovered by Gawronski et al. (2006) in the FIRST survey**
- **25 hrs JVLA time awarded to observe the sample at L- (1-2 GHz) and C-band (4-8 GHz)**
- **Observed in multiple configurations give ~1” resolution but also capture the diffuse emission**
- **First image now complete (left) thanks to Tessa Vernstrom (Dunlap Institute)**
- **Spectral analysis of the first source soon!**

Conclusions

- Machine learning provides a potentially powerful tool for identifying remnant radio galaxies and other sources in large radio surveys
- We cannot rely purely on one parameter (e.g. morphology), if we wish to achieve a robust candidate list and include more than a subset of remnants
- Machine learning can be computationally expensive and often requires specialist setups e.g. GPU rigs − plan for this!
- First results should be ready later this year, hopefully expanding the search to the full LoTSS survey (and beyond) as it progresses
- AGES-XL is now almost complete. Let me know if you would like to get involved!
- First images showing combined compact and extended emission of HYMOR produced − spectral analysis to follow shortly