Source-finding for SUMSS and MGPS-2

Tara Murphy

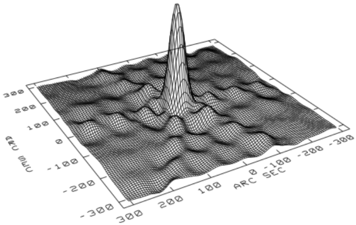
29th April 2010

What is source-finding?

1. Finding sources
   - Finding groups of bright pixels in images
   - Grouping pixels together in some way
2. Fitting and extracting sources
   - Modelling sources (e.g. with an elliptical Gaussian)
   - Measuring confidence of fit
3. Characterising sources
   - Measuring properties of sources
   - Classifying sources

Search And Destroy algorithms

1. Scan through each row of pixels
2. Find all pixels above a given cutoff
3. Merge all contiguous points above cutoff into islands
4. Generate initial estimate of strength and size of each island
5. Find least squares Gaussian fit to each island
6. Optional: If RMS residual is too high, try multiple Gaussian fit



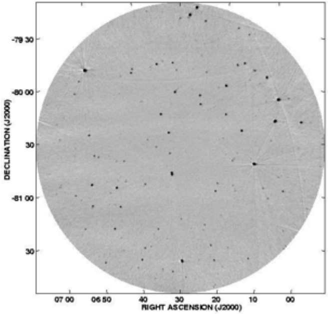
Ref: VSAD in AIPS, IMSAD in Miriad — Condon et al. 1998

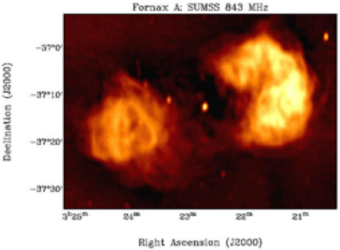
Pros and cons of Search and Destroy

- The main routines used are VSAD in AIPS, IMSAD in Miriad and SFIND (old-sfind) in Miriad
- Advantages of these algorithms:
  - Easy to understand
  - Easy to implement
  - Trivially parallel
- Disadvantages of these algorithms:
  - Very local focus – don't see global image properties
  - Only effective for compact (beam-like) sources
  - Provide limited characterisation
  - Not adaptive

The Sydney University Molonglo Sky Survey (SUMSS)

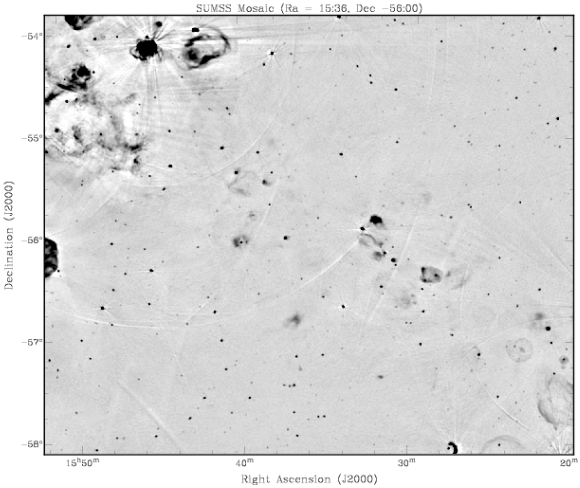
- 843 MHz survey of the Southern sky (to $\delta \leq -30$)
- Carried out from 1997 to 2007
- The Southern counterpart to NVSS
- 210 412 sources to a flux cutoff of 6 mJy beam^{-1}





- Mauch et al., 2003, MNRAS, 342, 1117

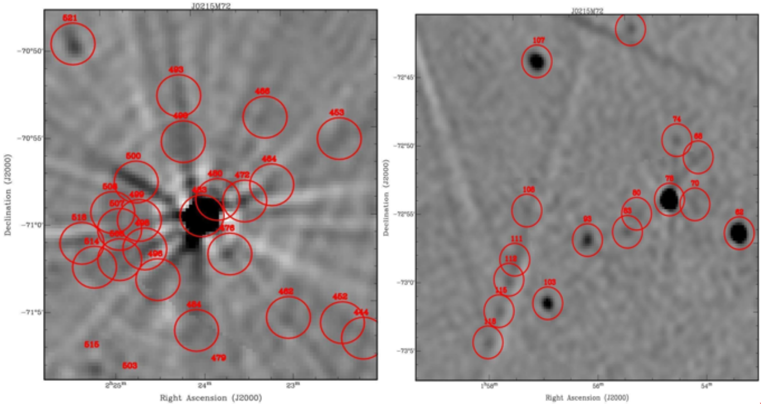
A typical SUMSS mosaic



Catalogue creation is a critical part of survey science

1. We are seeing **new** objects
   - We want to know what they are, what their properties are
   - Classification is an central tenant of observational science
   - Most basic classification (extended or compact) allows us to make different measurements
2. Not all objects found are 'real'
   - The telescope produces artefacts in the image
   - The source finder extracts these as sources
   - We don't want to include these in the catalogue

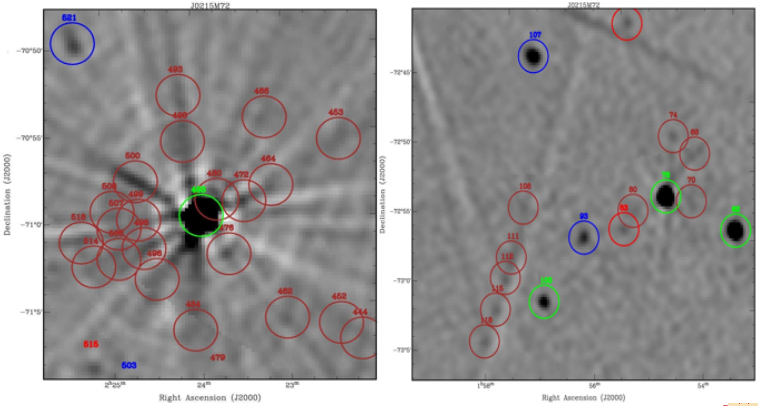
The source finder detects artefacts in the image



We used machine learning to classify each candidate source

- Decision trees are a group of algorithms that can learn a particular task
- Humans experts classify ∼ 1000's of objects which the decision tree uses to classify ∼ 100 000's of objects.
  1. It is *trained* on a small set of classified data
  2. It determines an optimal **tree** of decision points which best classify the data
  3. It is *applied* to the entire survey
- This avoids complex rule-based algorithms

97% of sources were classified correctly



Problems

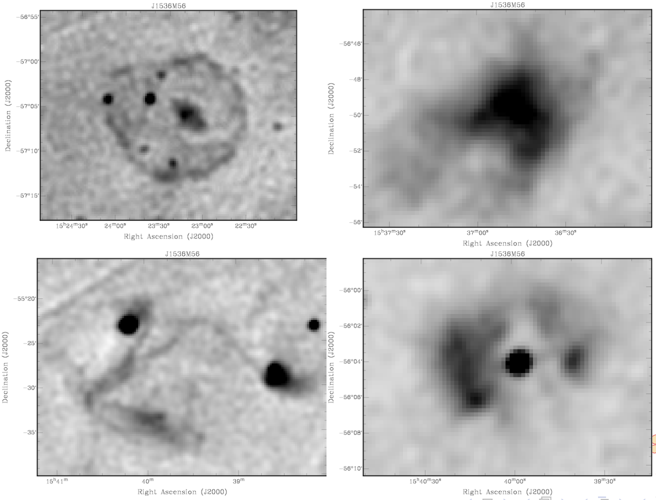
- This post-processing is still based on poor source finding
- Requires manual annotation of a large amount of data (could consider un-supervised machine learning methods)
- Doesn't solve the problem of extended source characterisation

The Molonglo Galactic Plane Survey (MGPS-2)

- 843 MHz survey of the southern Galactic plane (to $\delta \leq -30$)
- Galactic coverage $245^\circ < l < 365^\circ$, $|b| \leq 10^\circ$
- Carried out from 1997 to 2007
- Counterpart to SUMSS
- Restoring beam $45'' \times 45''\,\mathrm{csc}\,|\delta|$
- Good *uv*-coverage
- Some of the science goals are:
  - A radio study of planetary nebulae
  - To search for young SNR and high galactic latitude SNR
  - To search for transients between the two MGPS epochs
  - To search for pulsar wind nebulae and pulsar counterparts

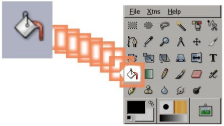
- Mauch et al., 2003, MNRAS, 342, 1117
- Murphy et al., 2007, MNRAS, 382, 382

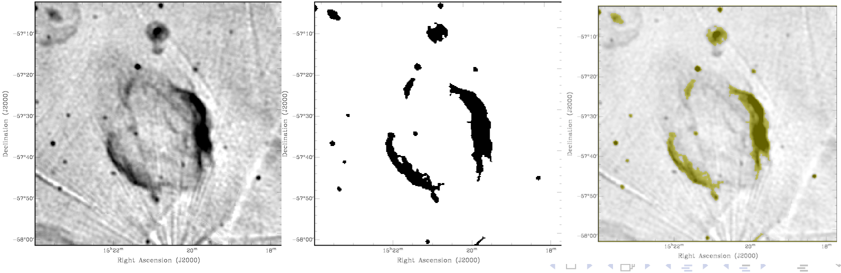
Galactic Plane sources are extremely diverse



We need different source finding algorithms

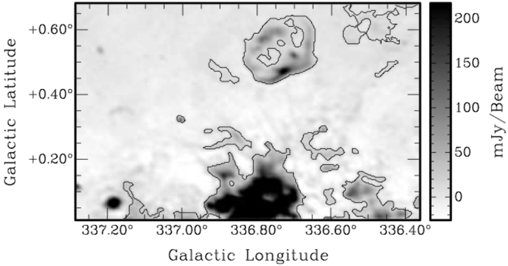
One example is the **floodfill** algorithm





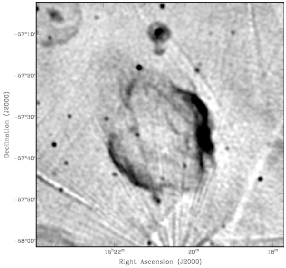
Floodfill + Search And Destroy

1. Run Search And Destroy and catalogue sources
2. Run floodfill and catalogue extended sources
3. For each SAD source identify corresponding floodfill source
4. Use this to identify compact and extended sources
5. Filter out artefacts and extended sources



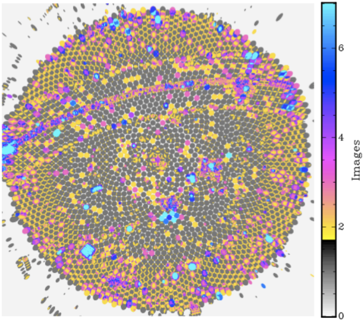
Problems

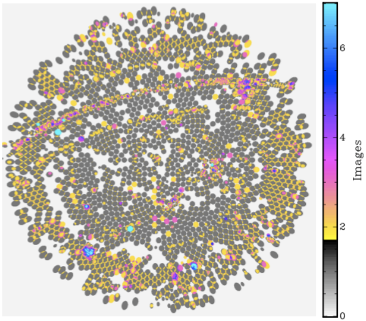
- How do we identify components of a single object?
- How do we (quantitatively) represent complex objects?



Searching for transients in the MOST archive

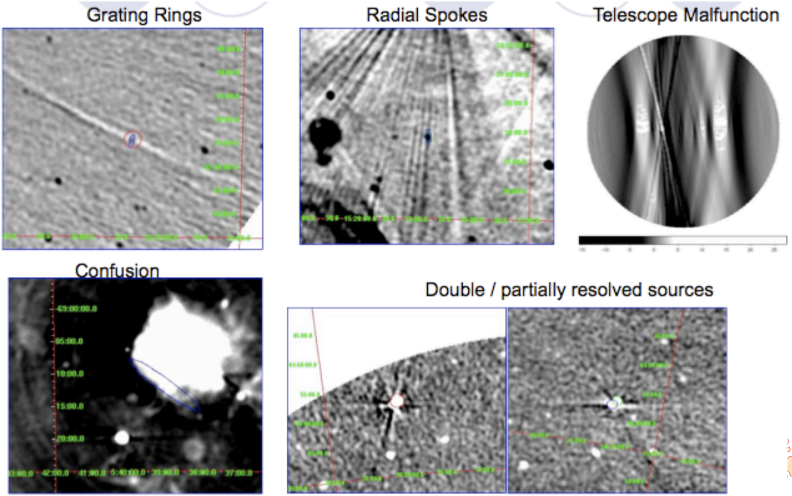
- Survey of 22 year Molonglo archive
- Lightcurves for 30,000 sources





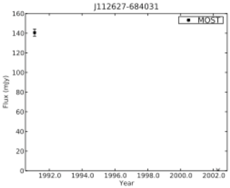
Bannister et al. 2010, in-prep-to-be-submitted-really-really-soon

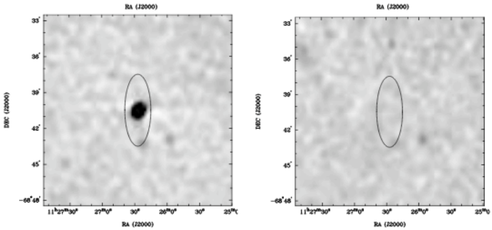
Artefacts are a significant problem for transient detection



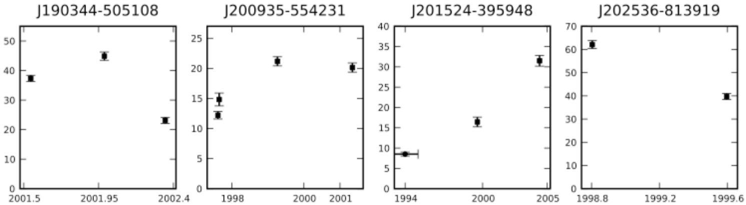
What is our completeness and reliability?

Transients (15):





Variables (55):



Problems

- Missing sources masquerade as ‘transients’
- Poor fits masquerade as ‘variables’
- To overcome many of the problems we used a 2-phase process
- We also did extensive manual inspection

Source-finding issues for ASKAP

- Completeness problems - missing sources (bug, algorithm?)
- Reliability problems
- Accurate characterisation of sources (e.g. flux, size)
- How to define cutoffs (noise level, peak)
- How to characterise extended and complex sources

- Problems for VAST:
  - Source finding errors appear as transients
  - Source measurement errors appear as variables
  - Getting accurate statistics is critical
  - Can't manually inspect everything