Source Detection in the Image Domain

Astro 6525, Fall 2015

Shami Chatterjee

Oct 2015

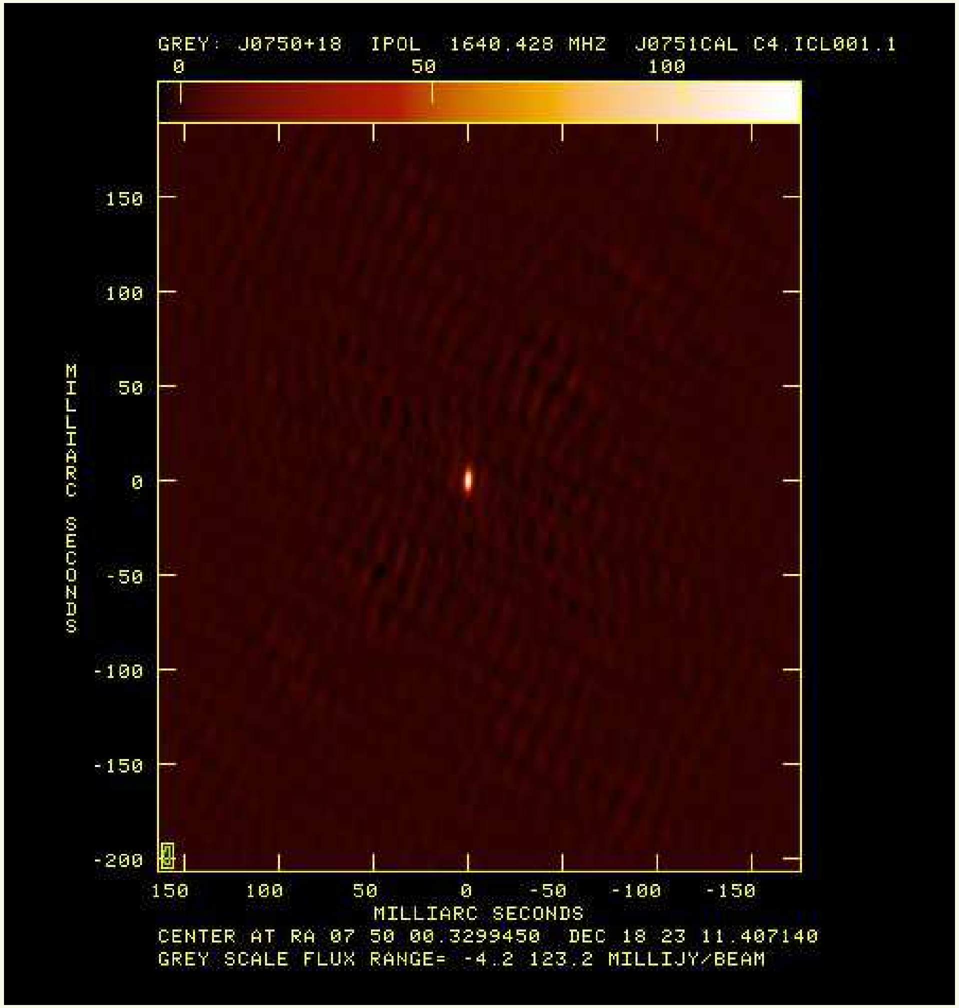
Source Identification

- Identify a source in the presence of noise.
- Localize it as best as possible given instrumental resolution.

Source Identification

- Identify a source in the presence of noise.
- Localize it as best as possible given instrumental resolution.

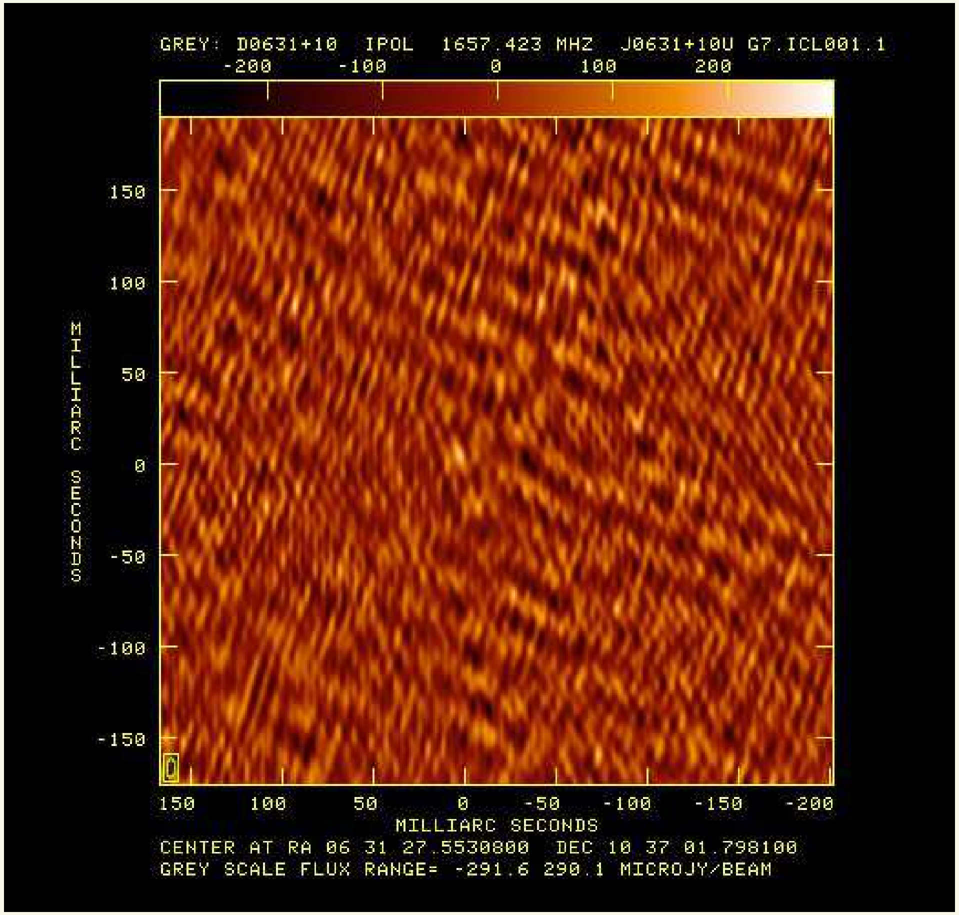
How well can we do? Depends on beam size, of course.



Source Identification

- Identify a source in the presence of noise.
- Localize it as best as possible given instrumental resolution.

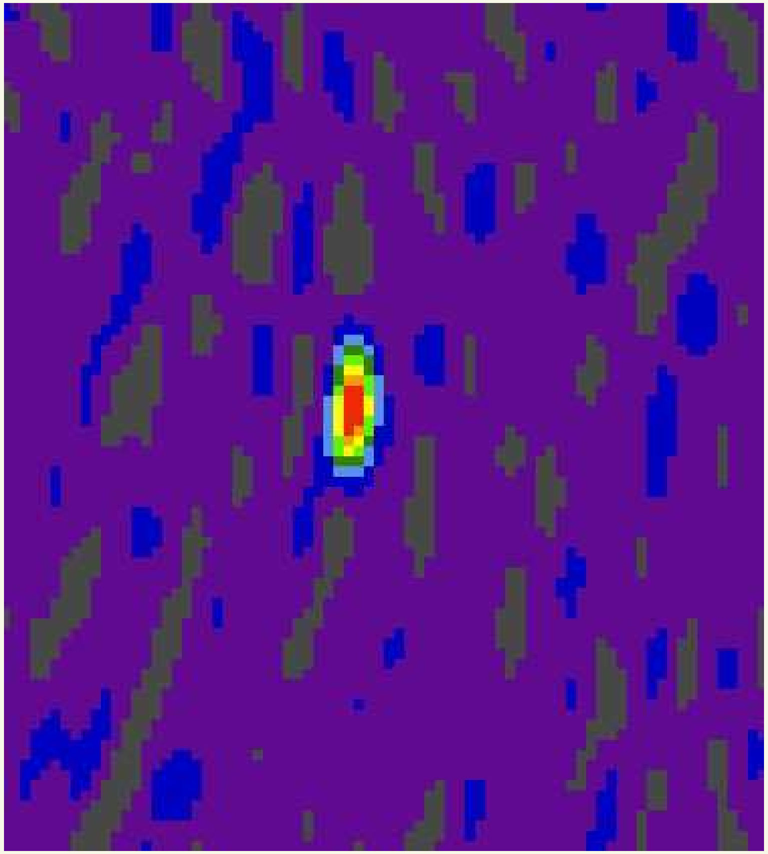
But it also depends on the detection signal-to-noise ratio...



Source Identification

- Identify a source in the presence of noise.
- Localize it as best as possible given instrumental resolution.

In fact, we can do better than the beam size
as long as we have enough signal-to-noise ratio.



Source Identification

- Identify a source in the presence of noise.
- Localize it as best as possible given instrumental resolution.

$\Rightarrow$ Resolution $\sim$ Beam Size / Signal-to-Noise ratio

(Of course, we're talking about localization *precision*, not localization *accuracy*.)

Source Detection Algorithms

- Identify "real" sources in the presence of noise in an image.
- Many different approaches to the problem, optimized for different image domains and properties of interest.

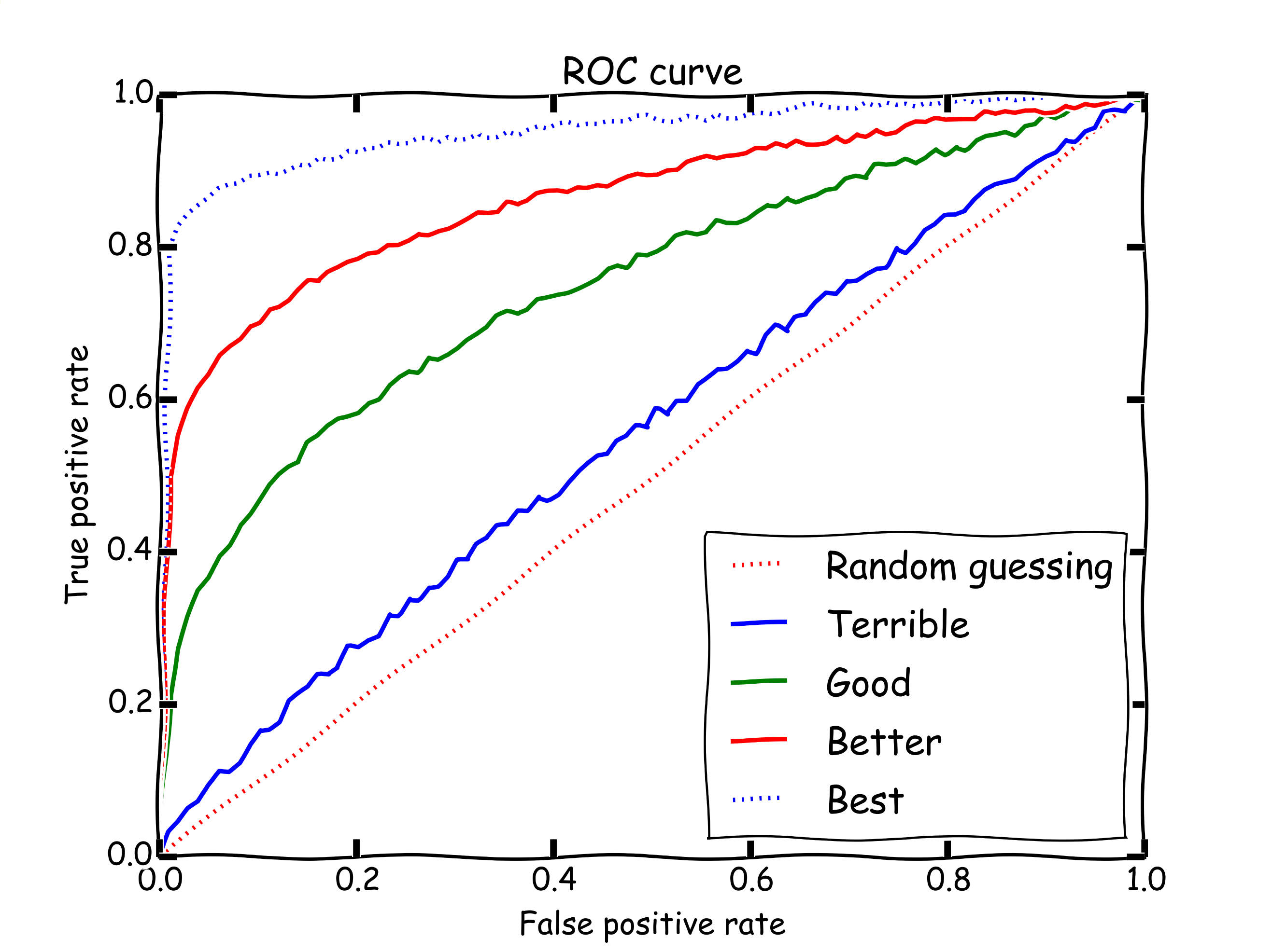
Source Detection Algorithms

- Identify “real” sources in the presence of noise in an image.
- Many different approaches to the problem, optimized for different image domains and properties of interest.
  - e.g., Optical, Radio, X-ray images?
    (Blurring from diffraction and seeing
    vs. Fourier transformed visibilities + CLEANing
    vs. Photon events (Poisson statistics) and sparse detector arrays)
  - Do we care about extended sources or want compact components?

Source Detection Algorithms

- Identify “real” sources in the presence of noise in an image.
- Many different approaches to the problem, optimized for different image domains and properties of interest.
  - e.g., Optical, Radio, X-ray images?
    (Blurring from diffraction and seeing
    vs. Fourier transformed visibilities + CLEANing
    vs. Photon events (Poisson statistics) and sparse detector arrays)
  - Do we care about extended sources or want compact components?
- Differrent trade-offs between speed and complexity,
  false positives and completeness.
- Examples: SFind, IMSAD, SExtractor, Aegean, BlobCat, etc. etc. etc.

Receiver Operating Characteristic (ROC) curves



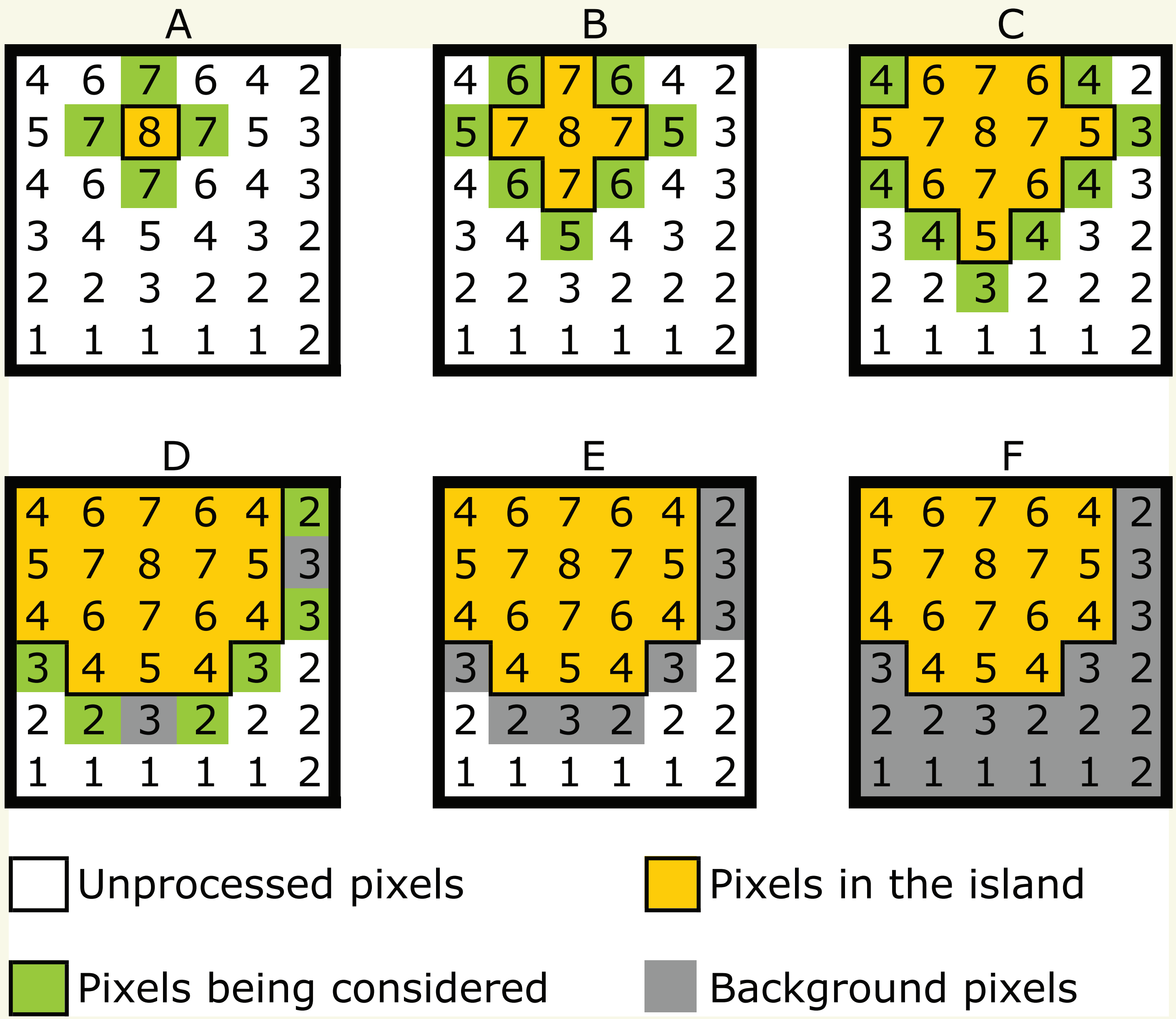
- ROC curve determines trade-off between false positive rate and completeness.
- Given ROC curve, choice of detection threshold determines completeness achieved and rate of false positives.
- This is up to the user to decide! Better to miss real signals, or to have fake detections? Very different answers depending on circumstances.

Algorithms

(Masias et al. 2012)

	Basic image transformations		
Herzog & Illingworth (1977)	Mean	Optical	SD
Le Fèvre et al. (1996)	Bijaoui	Multi-band	SD
Stetson (1987)	σ-clipping + Gaussian	N/A	SD
Slezak, Bijaoui & Mars (1988)	Gaussian + Bijaoui	Optical	SD
Bertin & Arnouts (1996)	σ-clipping	N/A	SD
Szalay, Connolly & Szokoly (1999)	Gaussian	Multi-band	FSD
Mighell (1999)	Mean	N/A	SD
Hopkins et al. (2002)	Gaussian	Radio	SD
Aptoula, Lefèvre & Collet (2006)	Morphological	Multi-band	SD
Yang, Li & Zhang (2008)	Median + Morphological	Optical	SD
Perret, Lefèvre & Collet (2008)	σ-clipping + Median + Morphological	Multi-band	SD
Haupt, Castro, & Nowak (2009)	Distilled sensing	Radio	SD
Lang et al. (2010)	Median	Multi-band	PSD
	Bayesian approaches		
Hobson & McLachlan (2003)	Markov-chain	N/A	SD
Savage & Oliver (2007)	Markov-chain	Infrared	SD
Feroz & Hobson (2008)	Nested sampling	N/A	SD
Carvalho, Rocha & Hobson (2009)	Multiple posterior maximization	Optical	SD
Guglielmetti, Fischer & Dose (2009)	Mixture model	X-ray	SD
	MF		
Irwin (1985)	Bijaoui + MF	Optical	SD
Vikhlinin et al. (1995)	σ-clipping + MF	X-ray	SD
Makovoz & Marleau (2006)	Median + MF	Multi-band	PSD
Melin, Bartlett & Delabrouille (2006)	Matched multi-filters	Radio and multi-band	PSD
Herranz et al. (2009)	Matched matrix filters	Radio	PSD
Torrent et al. (2010)	Boosting	Radio	FSD
	Multi-scale approaches		
Bijaoui & Rué (1995)	Wavelet	Optical	SD
Kaiser, Squires & Broadhurst (1995)	Mexican Hat	Multi-band	SD
Damiani et al. (1997)	Gaussian + Median + Mexican Hat	X-ray	SD
Starck et al. (1999)	Wavelet	Mid-infrared	FSD
Lazzati et al. (1999)	σ-clipping	X-ray	SD
Freeman et al. (2002)	Mean + Mexican Hat	X-ray	SD
Starck (2002)	Wavelet + Ridgelet	Infrared	SD
Starck, Donoho & Candès (2003)	Wavelet + Curvelet	Infrared	SD
Vielva et al. (2003)	Mexican Hat (spherical)	Radio	PSD
Belbachir & Goebel (2005)	Contourlet + Wavelet	Infrared	FSD
Bijaoui et al. (2005)	Wavelet + PSF smoothing	Multi-band	SD
González-Nuevo et al. (2006)	Mexican Hat (family)	Radio	PSD
Starck et al. (2009)	Multi-scale variance stabilization	γ-ray	SD
Peracaula et al. (2009b)	Gaussian + Wavelet	Radio	PSD
Peracaula et al. (2011)	Gaussian + Wavelet	Radio and infrared	ESD
Broos et al. (2010)	Wavelet	X-ray	SD

Source Identification: Flood Fill

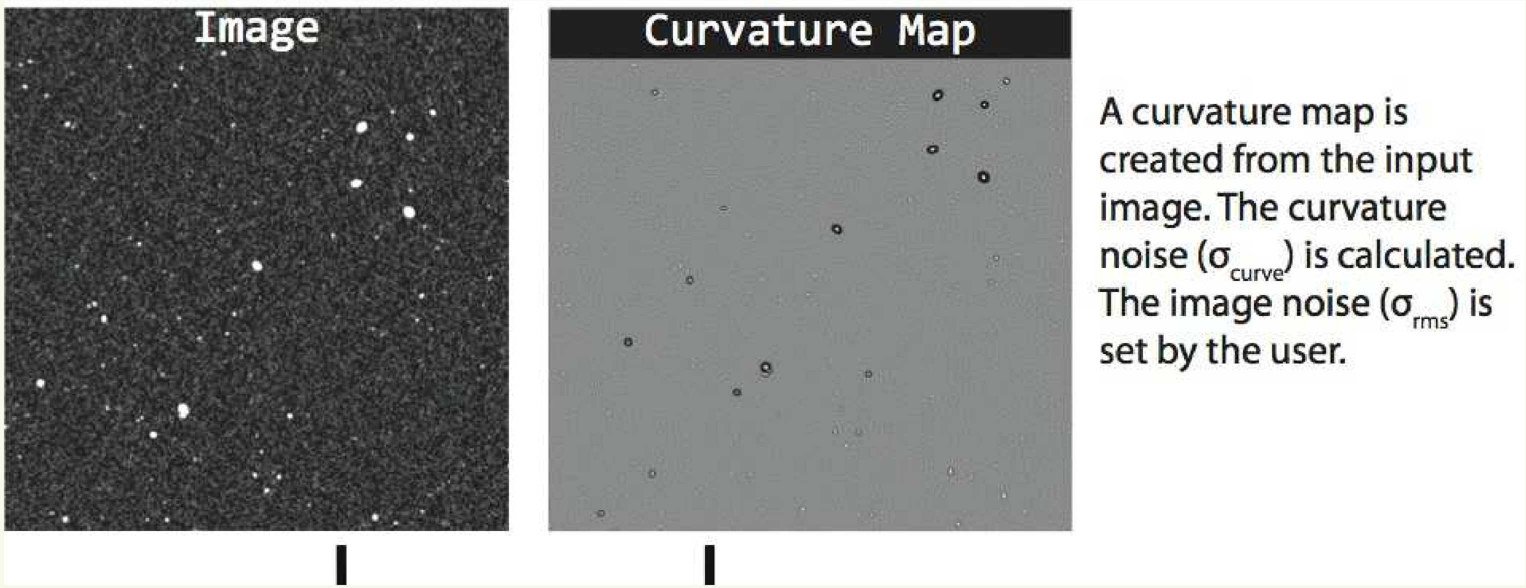


Source detection with Aegean

Aegean: Part of Paul Hancock's thesis work with Murphy, Gaensler, et al.
(Hancock et al. 2012, MNRAS, 422, 1812)

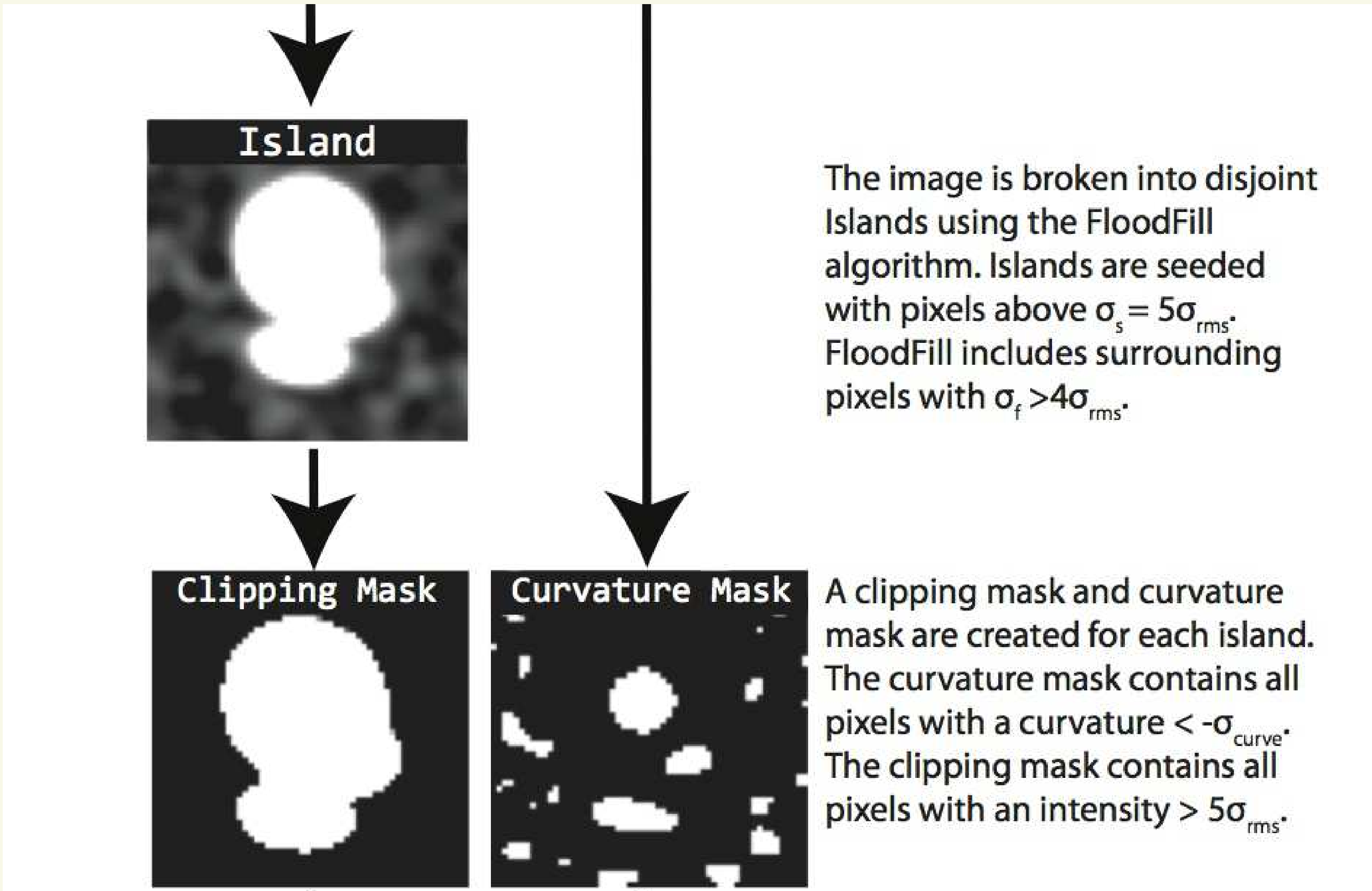
Source detection with Aegean

Aegean: Part of Paul Hancock's thesis work with Murphy, Gaensler, et al. (Hancock et al. 2012, MNRAS, 422, 1812)



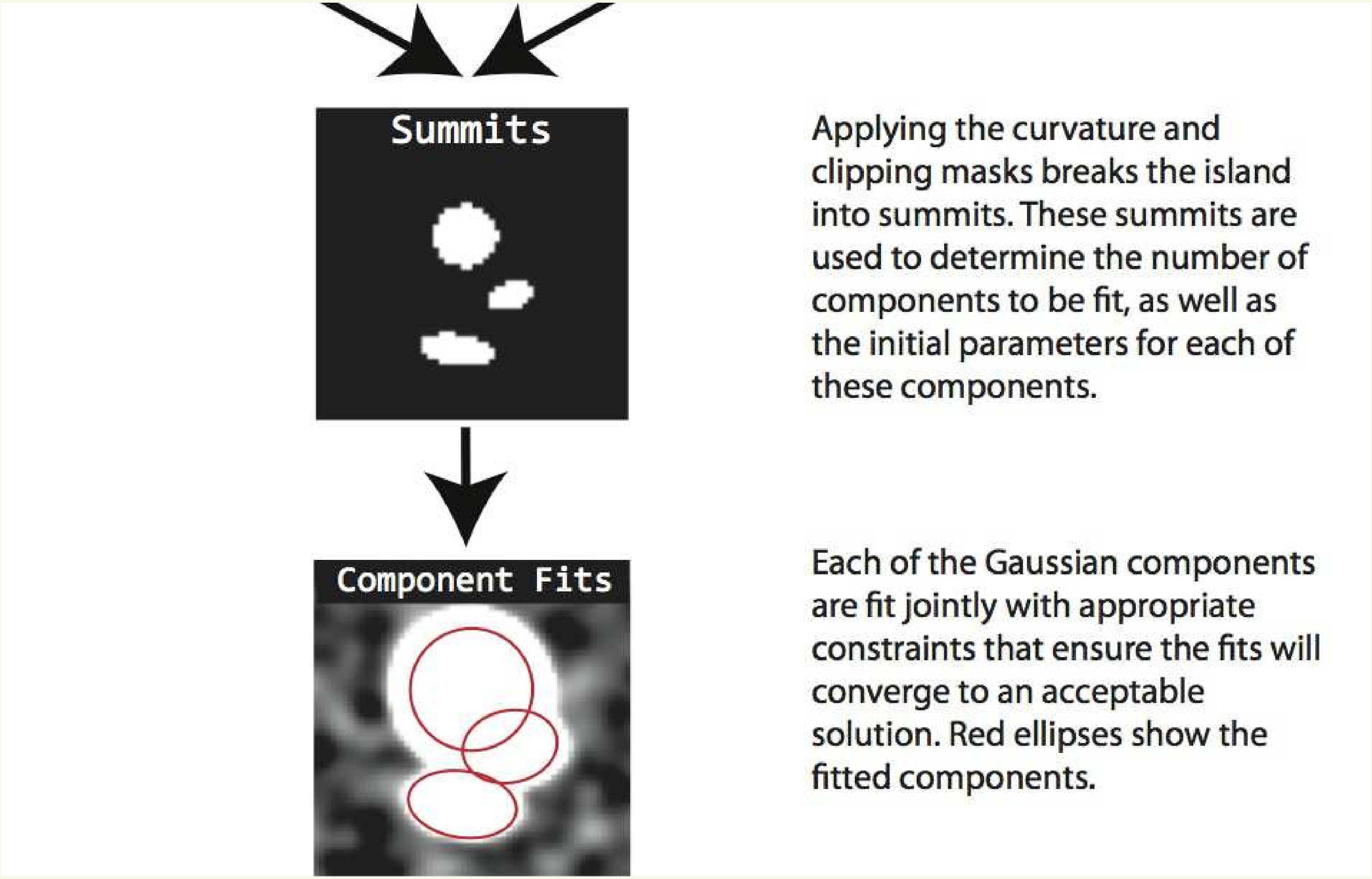
Source detection with Aegean

Aegean: Part of Paul Hancock's thesis work with Murphy, Gaensler, et al.
(Hancock et al. 2012, MNRAS, 422, 1812)

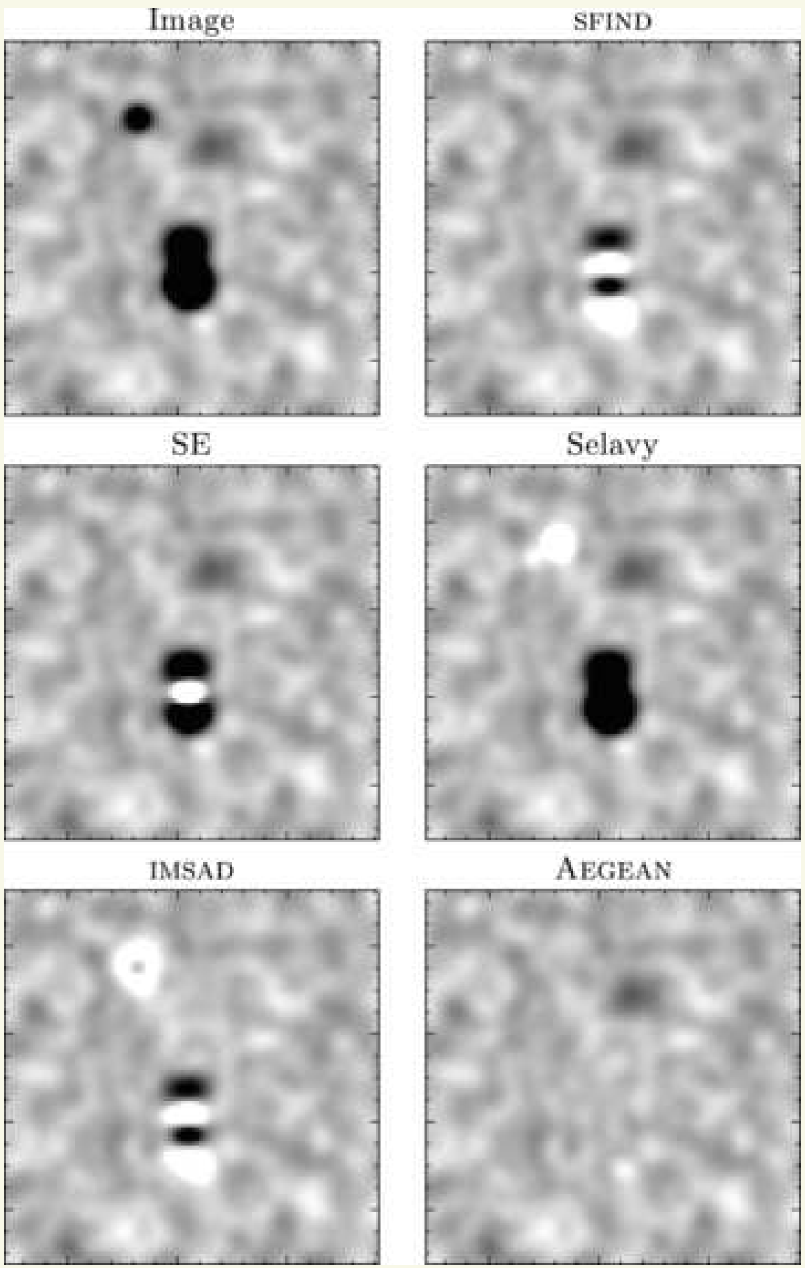


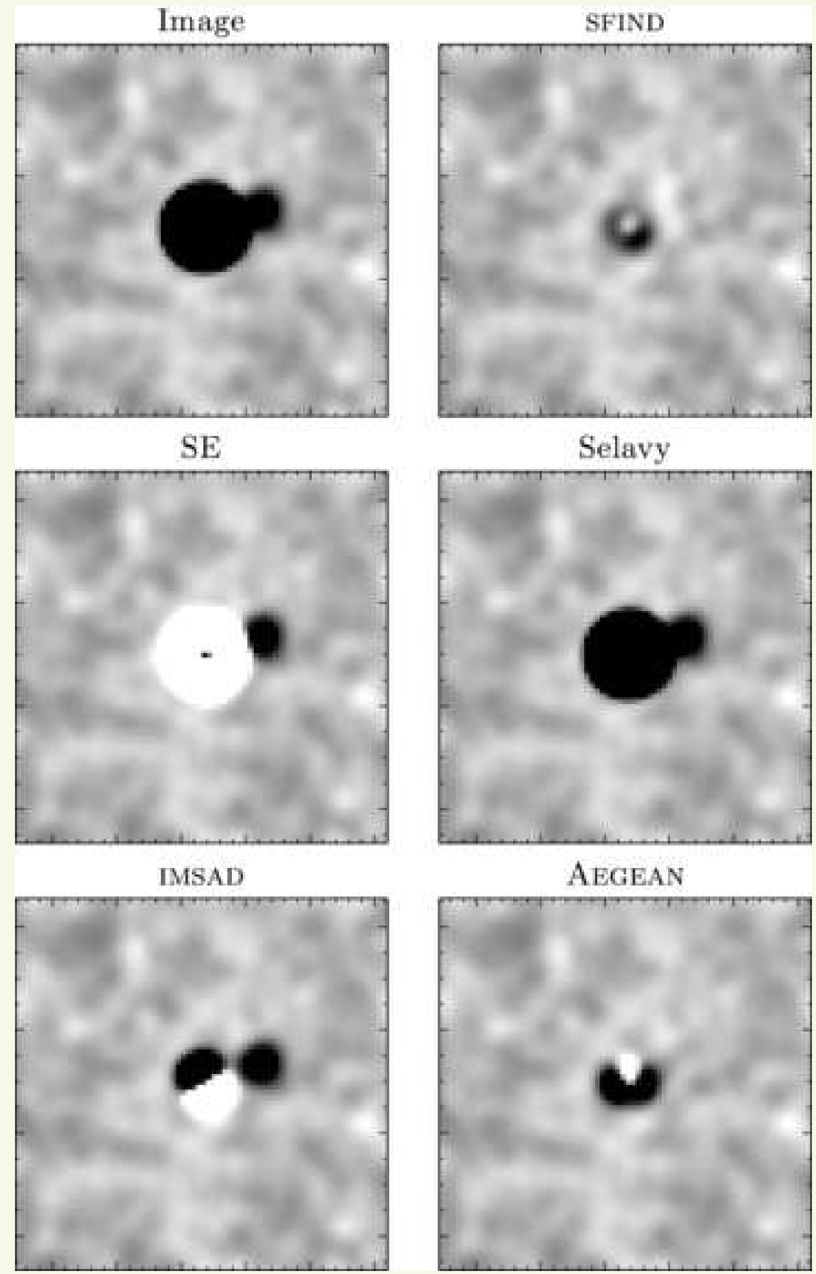
Source detection with Aegean

Aegean: Part of Paul Hancock's thesis work with Murphy, Gaensler, et al. (Hancock et al. 2012, MNRAS, 422, 1812)



Source detection with Aegean: Performance Comparisons

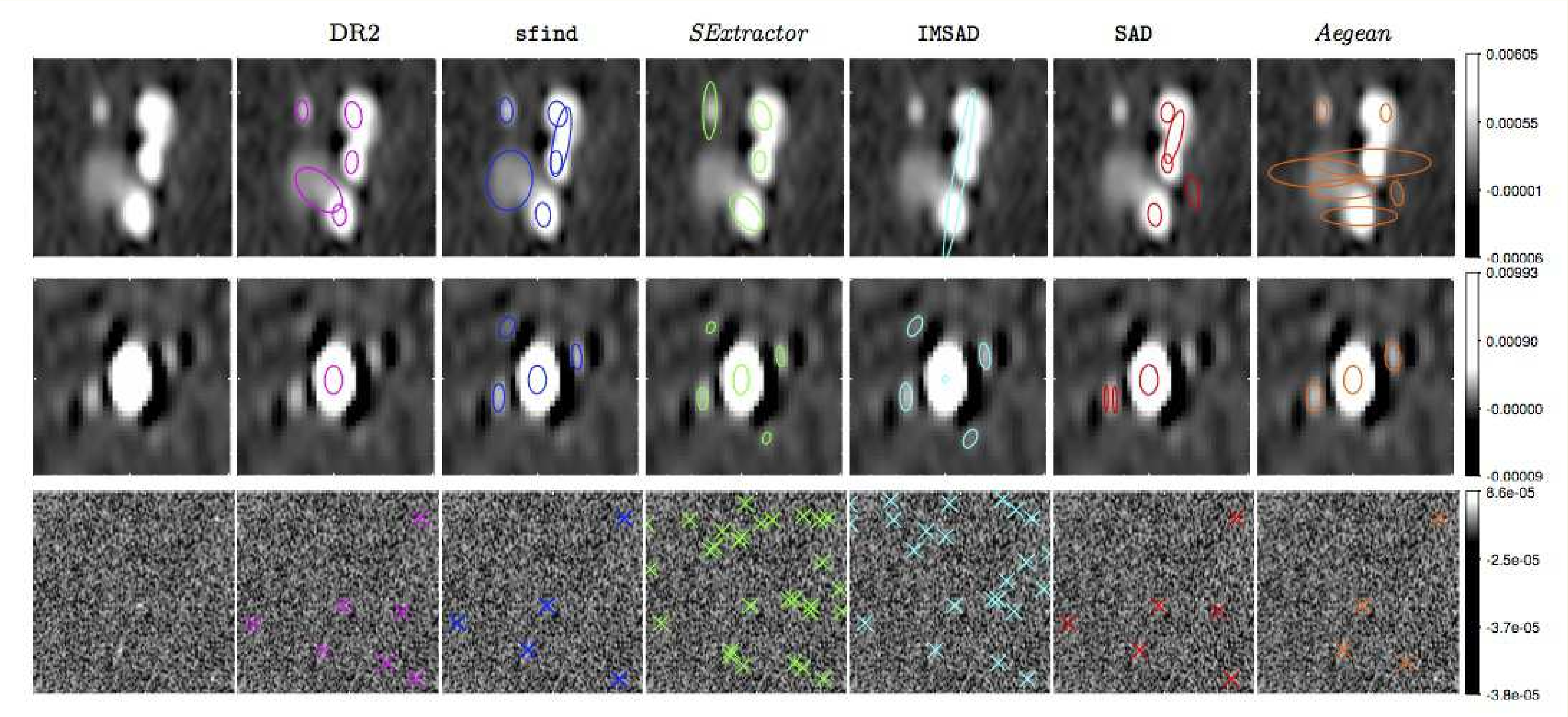




Source Finding: Real World Performance

- Source finders compared on real VLA observations

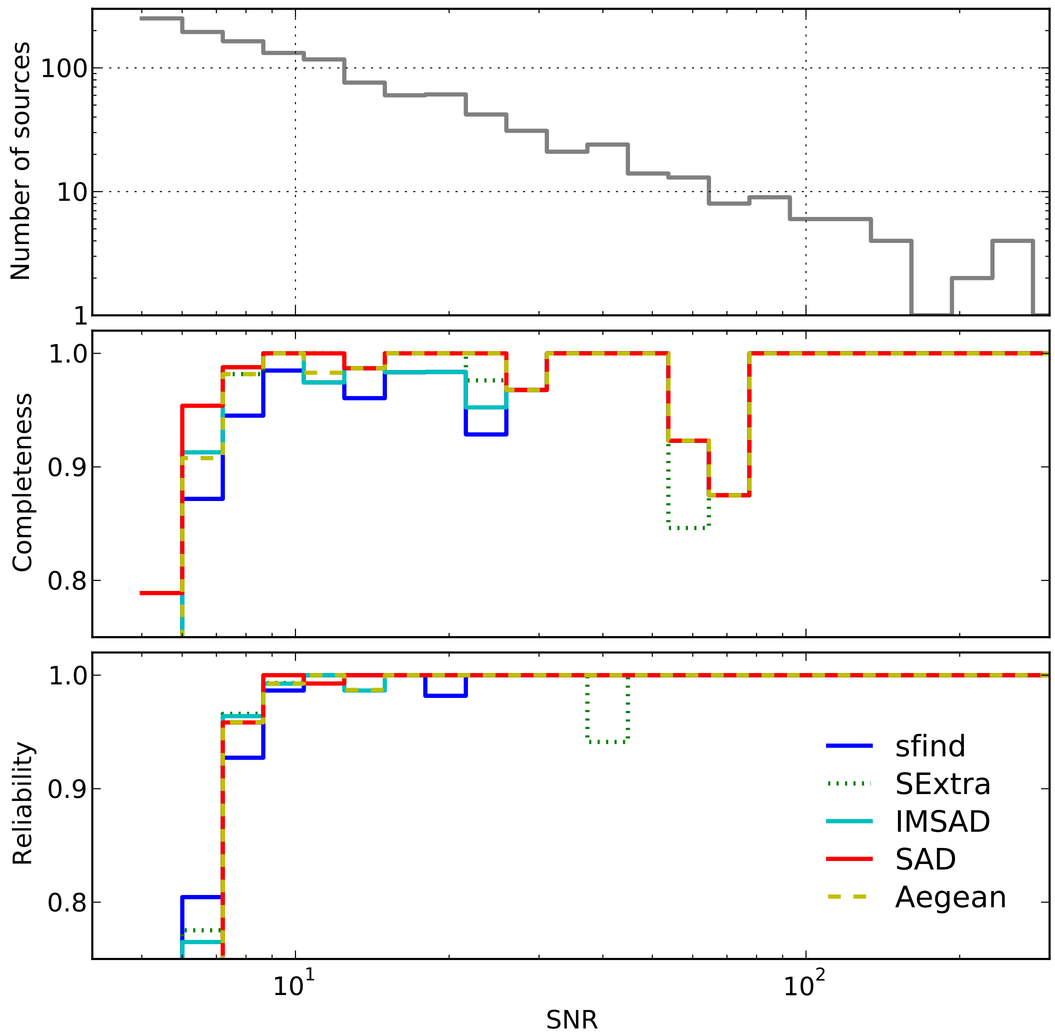
(Mooley et al. 2013, arXiv:1303.6282)



Comparison of source finder performance vs DR2 catalog for
Top: a blended source,
Mid: a source with sidelobes,
Bottom: a wide field area.

For sources in each S/N ratio bin (top), Mooley et al. compare completeness (middle) and reliability (bottom) of source finders.

(Highly blended sources cause dip at SNR$\sim$70)

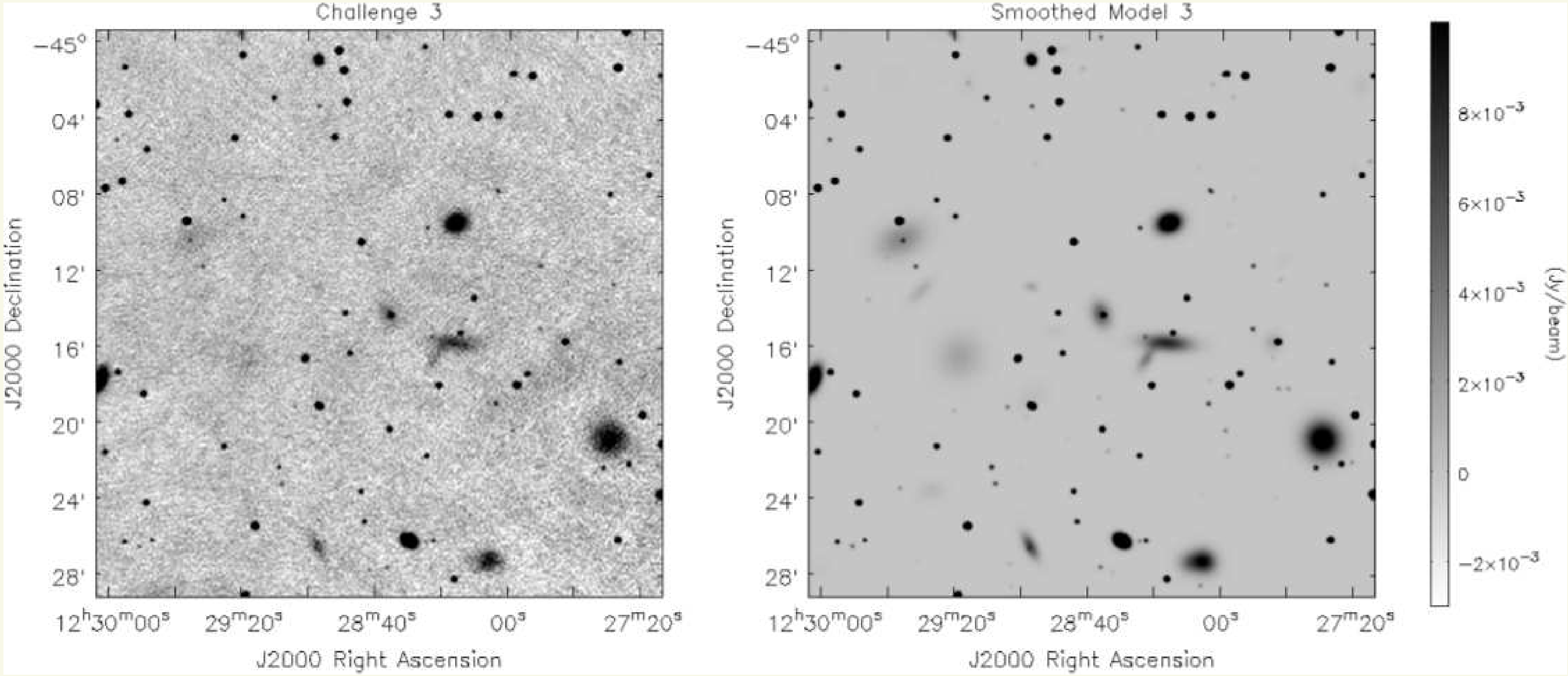


Source Finding: Simulated Performance

The ASKAP/EMU Source Finding Data Challenge
Hopkins et al. 2015; arXiv:1509.03931

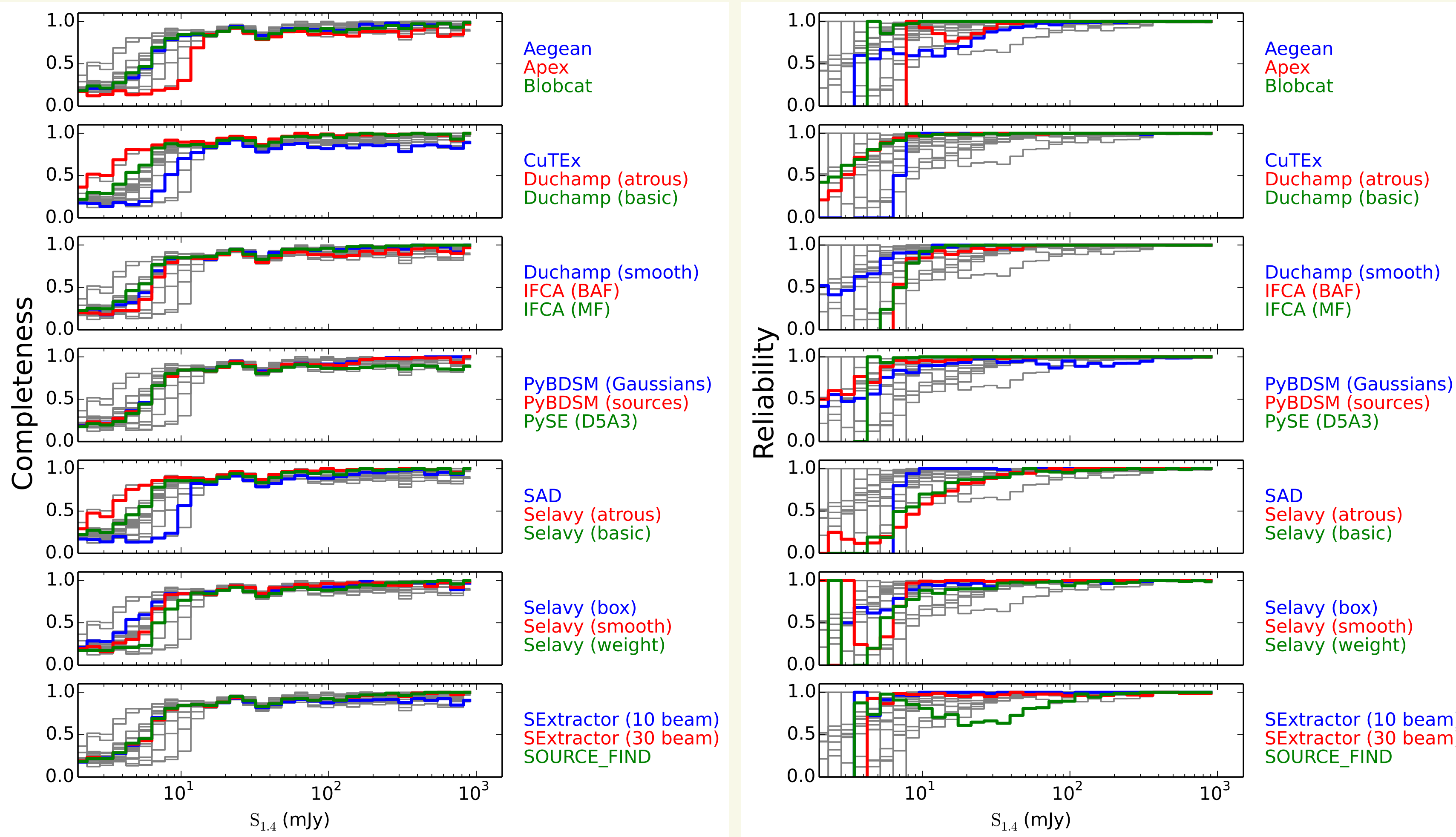
*The Evolutionary Map of the Universe (EMU) is a proposed radio continuum survey of the Southern Hemisphere up to declination $+30^\circ$, with the Australian Square Kilometre Array Pathfinder (ASKAP). EMU will use an automated source identification and measurement approach that is demonstrably **optimal**, to maximise the reliability, utility and robustness of the resulting radio source catalogues. As part of the process of achieving this aim, a "Data Challenge" has been conducted, providing international teams the opportunity to test a variety of source finders on a set of simulated images. The aim is to quantify the accuracy of existing automated source finding and measurement approaches, and to identify potential limitations. The Challenge attracted nine independent teams, who tested eleven different source finding tools. [...] Most finders demonstrate completeness levels close to 100% at $\approx 10\,\sigma$ dropping to levels around 10% by $\approx 5\,\sigma$. The reliability is typically close to 100% at $\approx 10\,\sigma$, with performance to lower sensitivities varying greatly between finders. All finders demonstrate the **usual trade-off between completeness and reliability**, whereby maintaining a high completeness at low signal-to-noise comes at the expense of reduced reliability, and vice-versa. [...]*

Source Finding: Simulated Performance



A subsection of the third Data Challenge image (left), and the input source distribution to this image (right). 20% of the sources are assigned a non-negligible physical extent. The extended sources are modelled as two-dimensional elliptical Gaussians.

Source Finding: Simulated Performance



The completeness and reliability fractions (left and right respectively) as a function of input source flux density (completeness) or measured source flux density (reliability) for each of the tested finders for Challenge 3. The grey lines show the distribution for all finders in each panel, to aid comparison for any given finder.

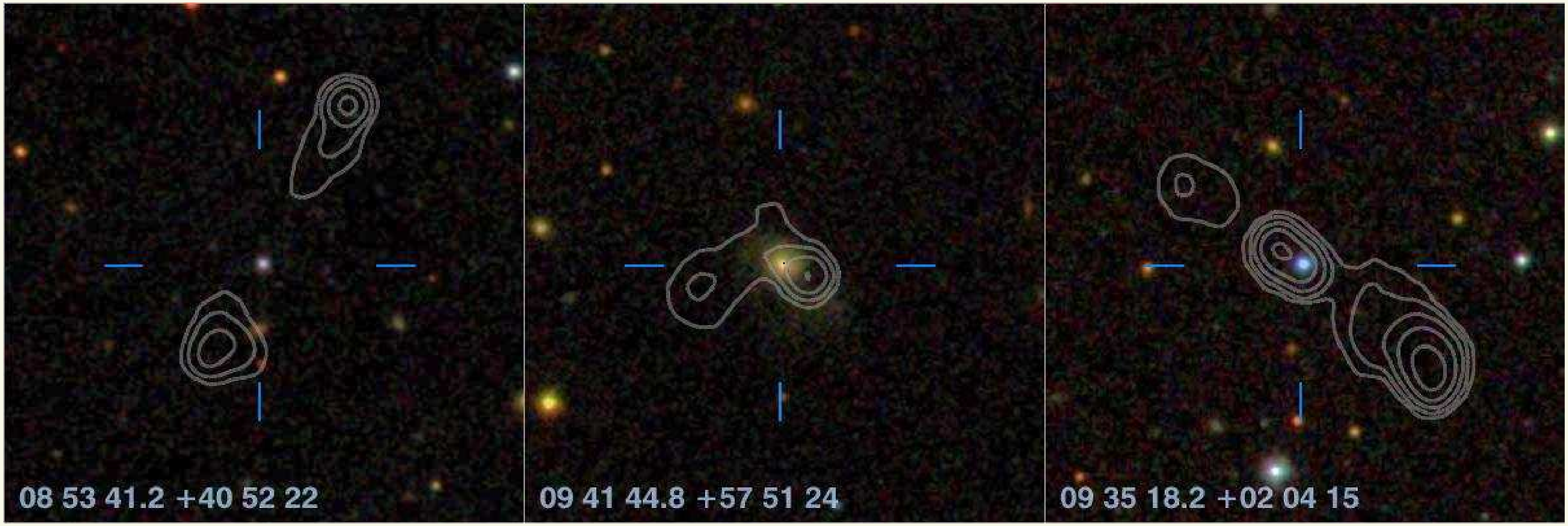
Bottom line for Source Detection Algorithms

- Source finders are typically optimized for different domains.
- Trade-off between completeness and reliability (false positives).
- All source finders are pretty good.
  No source finder is perfect.
  If you're using one, understand its strengths and weaknesses first.
- There's room to do better!

But what is a “source”?

- Radio synthesis imaging: usually groups of CLEAN components.

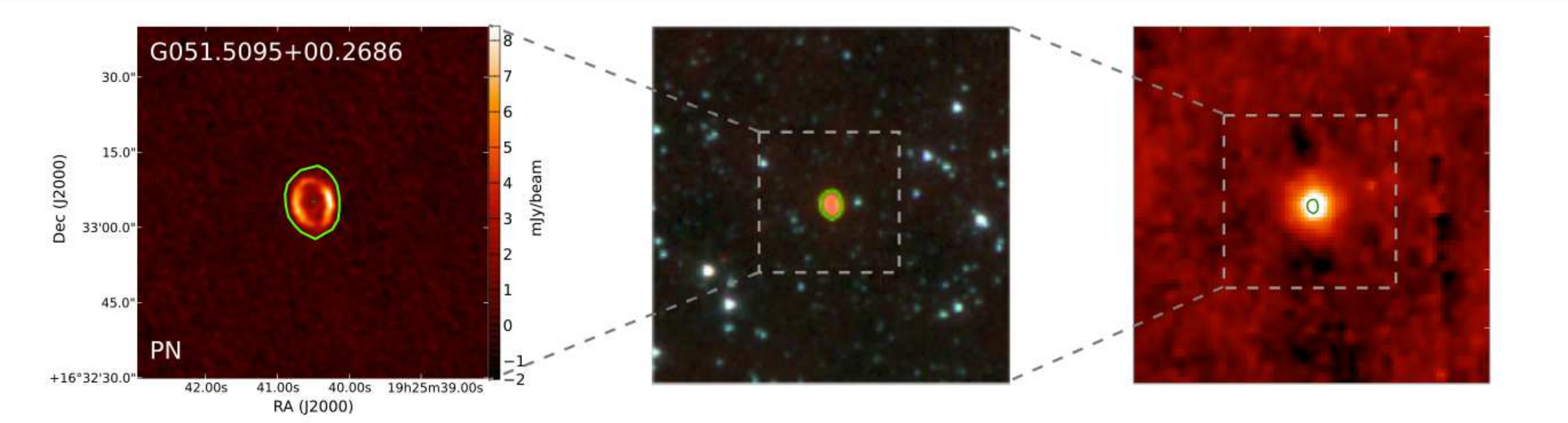
→ A physical source might span multiple such groups.
→ Multiwavelength cross-match may not even overlap detected source.



⇒ Problematic if you want to do science with your source catalog.

“Band Merging”

- Even at nearby wavelengths, change in telescope resolution λ/D:
$\Rightarrow$ different apparent morphology.

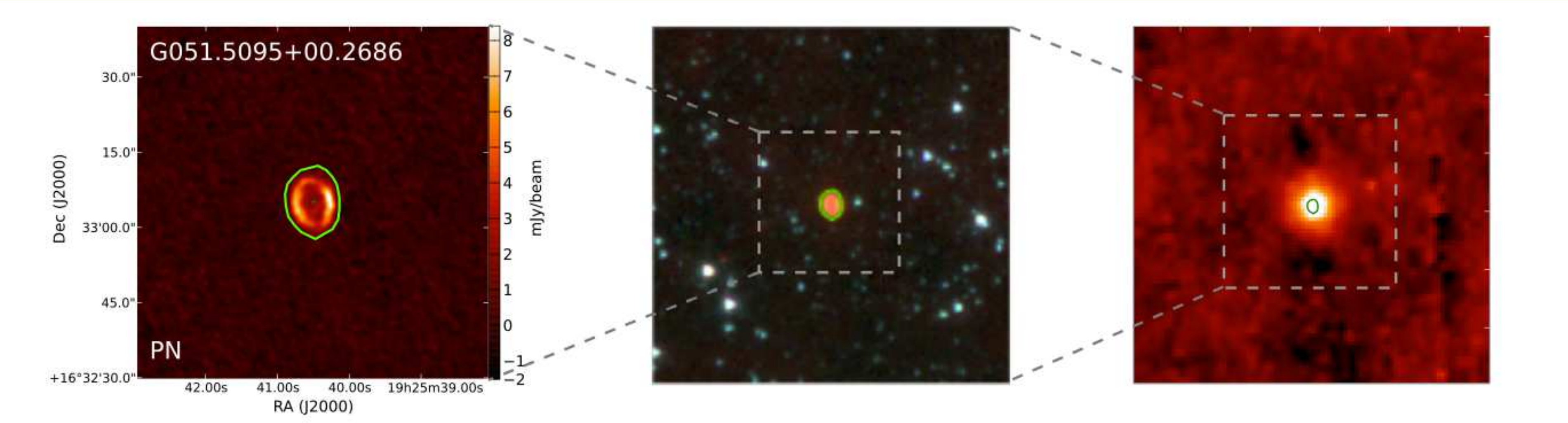


$\Rightarrow$ Merging information across different wavelength bands (“band merging”) is tricky when operating on large scale surveys.

“Band Merging”

- Even at nearby wavelengths, change in telescope resolution λ/D:

$\Rightarrow$ different apparent morphology.



$\Rightarrow$ Merging information across different wavelength bands (“band merging”) is tricky when operating on large scale surveys.

How to proceed? Use prior information ...

THE ASTROPHYSICAL JOURNAL, 679:301–309, 2008 May 20

PROBABILISTIC CROSS-IDENTIFICATION OF ASTRONOMICAL SOURCES

TAMÁS BUDAVÁRI AND ALEXANDER S. SZALAY
Department of Physics and Astronomy, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218;
and Max-Planck-Institute für Astrophysik, Karl-Schwarzschild-Strasse 1, 85748 Garching, Germany
Received 2007 July 11; accepted 2008 February 9

ABSTRACT

We present a general probabilistic formalism for cross-identifying astronomical point sources in multiple observations. Our Bayesian approach, symmetric in all observations, is the foundation of a unified framework for object matching, where not only spatial information, but also physical properties, such as colors, redshift, and luminosity, can be considered in a natural way. We provide a practical recipe to implement an efficient recursive algorithm to evaluate the Bayes factor over a set of catalogs with known circular errors in positions. This new methodology is crucial for studies leveraging the synergy of today's multiwavelength observations and to enter the time domain science of the upcoming survey telescopes.

Subject headings: astrometry — catalogs — galaxies: statistics — methods: statistical

Mechanical Systems and Signal Processing 66-67 (2016) 120–136

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Bayesian source identification using local priors

M. Aucejo *, O. De Smet

Structural Mechanics and Coupled Systems Laboratory, Conservatoire National des Arts et Métiers, 2 Rue Conté, 75003 Paris, France

A R T I C L E I N F O

Article history:
Received 4 July 2014
Received in revised form
16 January 2015
Accepted 1 May 2015
Available online 23 May 2015

Keywords:
Inverse problem
Force identification
Group generalized Tikhonov regularization
Local priors

A B S T R A C T

This paper is concerned with the development of a general methodology for identifying mechanical sources from prior local information on both their nature and location over the studied structure. For this purpose, the formulation of the identification problem is derived from the Bayesian statistics, that provides a flexible way to account for local a priori on the distribution of sources. Practically, the resulting optimization problem can be seen as a group generalized Tikhonov regularization, that is solved in an iterative manner. The main features of the proposed identification method are illustrated with both numerical and experimental examples. In particular, it is shown that properly exploiting the local spatial information drastically improves the quality of the source identification.

The VLA Sky Survey

Source finding and matching: particular concern for large sky surveys.
$\rightarrow$ LSST will image all optical sky in the time domain.
$\rightarrow$ SKA pathfinders like ASKAP: designed to maximize survey speed metric.

The VLA Sky Survey

Source finding and matching: particular concern for large sky surveys.
$\rightarrow$ LSST will image all optical sky in the time domain.
$\rightarrow$ SKA pathfinders like ASKAP: designed to maximize survey speed metric.

VLA Sky Survey:

- EVLA upgrade $\rightarrow$ 10$\times$ spectral coverage, sensitivity, versatility. Need radio atlas successor to NVSS, FIRST.
- All-sky survey: $\delta > -40^\circ$; 33,885 sq deg.
- B-array; 2–4 GHz; 2.4” resolution.
- On-the-fly mapping; sensitivity $\sim$70 μJy.
- Three epochs: transients, polarization science, etc.

Now being planned - see, e.g.,
`https://science.nrao.edu/science/surveys/vlass`