cea
France

Testing Source Finders with Simulated Source Maps

Hugh Garsden

Stéphane Corbel

+ WG (John, Dario, Antonia, Alexander etc.).

The Source Finders

PyBDSM	Pyse	Aegean
• Python interface, C/ Fortran • In use in LOFAR Pipeline (MSSS) • Lots of functionality including wavelets (extended sources), parallel threads • Under continual development • David Rafferty and Niruj Mohan (Leiden)	• Python interface, C • Intended(?) for LOFAR TraP • Less functionality • Compact, faster than PyBDSM • Maintained at Amsterdam • Hanno Spreeuw (Amsterdam)	• Python interface, C • New • Modern island flood-fill algorithm, parallelism • Under continual development • Paul Hancock (Sydney) Leave for now

cea
France

Source Maps

- Correlated noise (clean or dirty beam)
- Random elliptical Gaussians

- Benefits
  - Volume (statistics)
  - Control

Test drivers

- Python scripts built for 3 source finders
- Batch run thousands of maps, real or simulated
- Match against catalogs, real or simulated
- Vary properties of simulated maps, e.g.. blending , noise, size, numbers
- Vary parameters of source finders

The Results

- The source finders work!
    - 99% hit rate on easy maps (10000 sources)
    - 1024px map in 2-3 secs
        - Pyse faster
    - Locations very accurate
    - Flux pretty accurate (2%)
- When they don't work so well:
    - Not tuned properly (later)
    - blended sources, 80%
    - Extended sources
    - Can be outliers

Issue: Parameters

- Questions
  - PyBDSM has 50 input parameters
  - Pyse has "—detection" "—analysis" threshold parameters
  - Set the wrong values, things quickly go bad
    - I and others obtained values by discussion, experiment, knowledge of map properties
- Need "set and forget" option
- Answer: turn on False Detection Rate algorithm

UnivEarthS

Issue: False Detection Rate

	PyBDSM	Pyse
Without FDR, parameters by experiment	Hit Rate 99.3% False positives 0.0043%	Hit Rate 99.8% False Positives 0.16%
With FDR of 5%	Hit rate 99.8% False positives 544% **OR** Hit rate 68% False positives 0%	Hit rate 99.9% False Positives 5.64%

- The point: Seems like there are times FDR shouldn't be used
- Need a "set and forget" to control the FDR
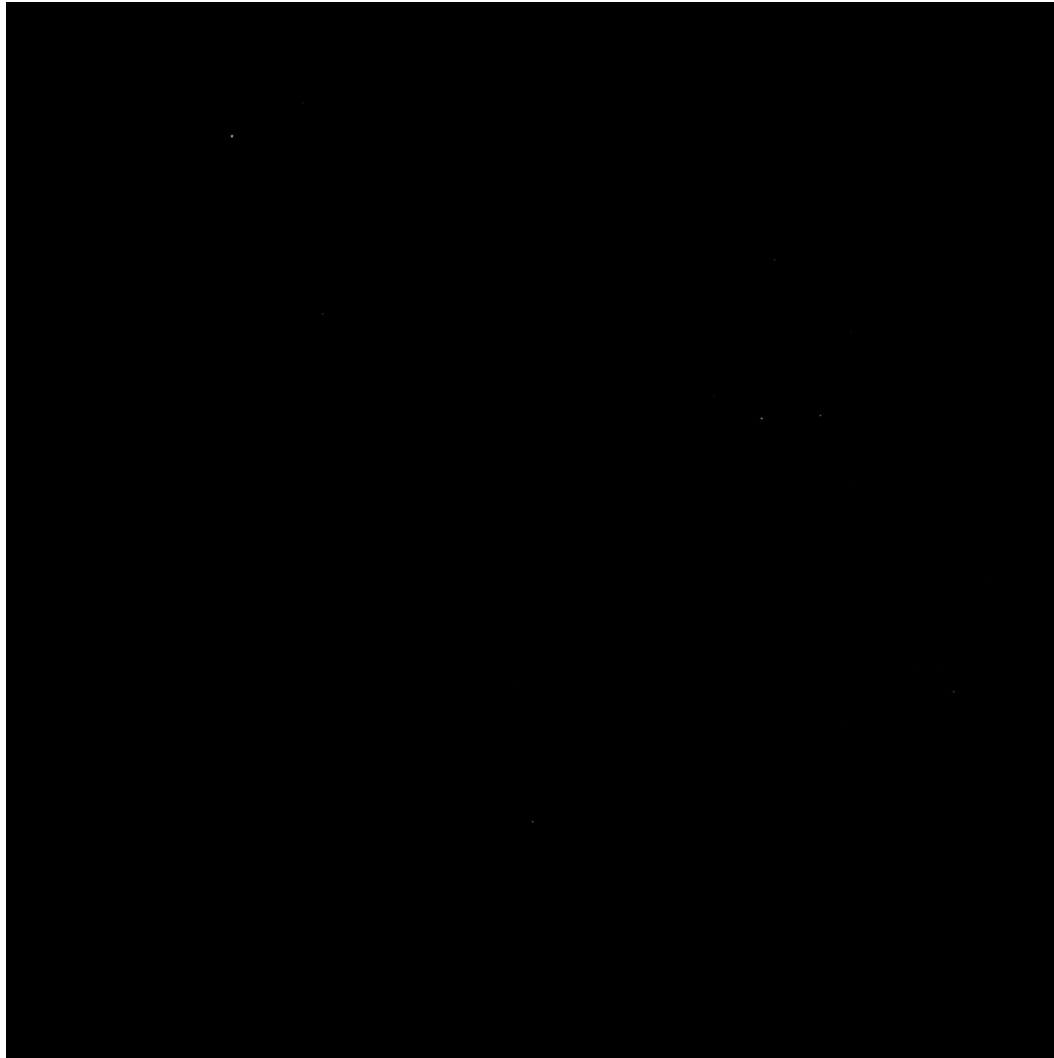
Future

- More realistic maps
  - Chiara, Dario, others
  - Database of maps we can all use
  - Simulate sources in a measurement set
  - CLEAN it, generate map

Hancock Maps

4800px
10000 sources

Where? They are
faint.

Future

- ASKAP EMU Source Finder Challenge (simulated maps)

- Work on false detection rate/ parameters

- Recode Pyse in C++, parallelize, GPU

- Follow development of Aegean

- PyBDSM continually being improved

France

UnivEarthS