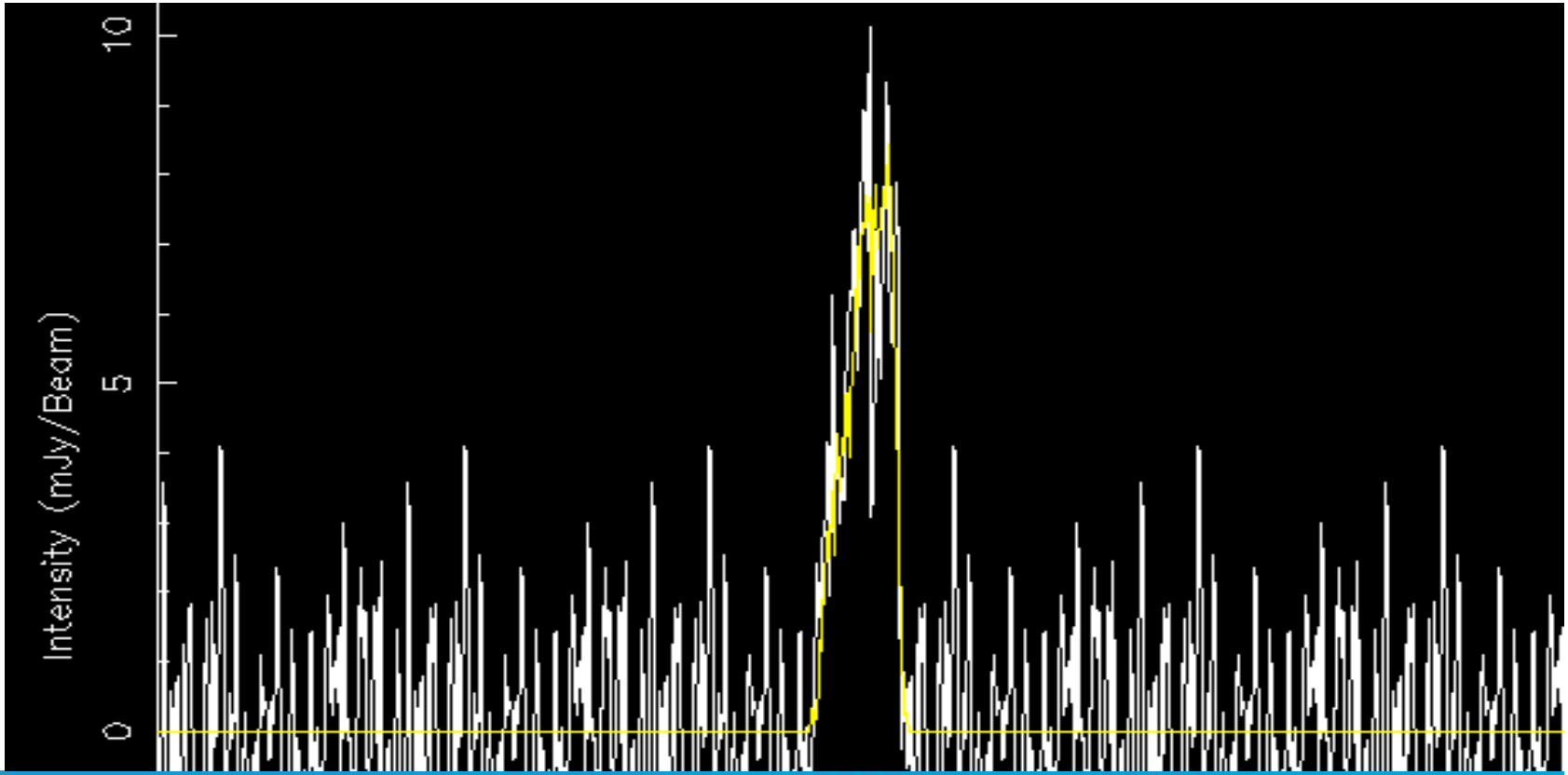


www.csiro.au

HI source finding: 2 novel approaches

ATNF, December 2010

CSIRO

Outline

- Background
  - What is source finding?
  - How do we assess source finders?
- Motivation for developing new HI source finding methods
- Novel source finding methods
  - General idea
  - Iterative 3-D median smoothing
  - Characterised noise method
  - Wavelet method
- Summary

CSIRO

Background

What is source finding?

- Traditionally, S'Finding has 6 components
  - Pre-processing
  - Finding source voxels/pixels
  - Recovering extent of each source
  - Merging source components
  - Characterising sources
  - Filtering sources
- Only 1 component is actually source finding!!!
- Other components aid/make use of actual source finding

CSIRO

How do we assess source finders?

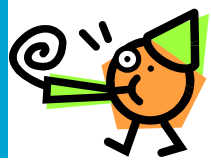
- Standard criterion
  - Reliability = probability of a source being real
  - Completeness = fraction of all sources recovered
  - Source parameters = how accurate are they?
  - This tests end-to-end performance of all 6 components.
- Alternatively
  - Only completeness is related to source finder performance
  - Explicitly use `refined' reliability and completeness
  - Bias. Are some sources more likely to be detected?

CSIRO

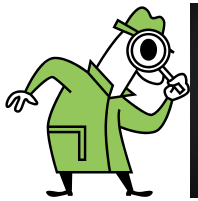
Motivation

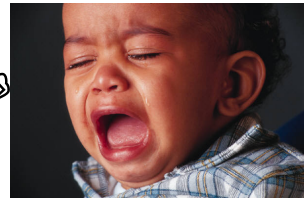
Motivation

- ASKAP = high resolution HI data cubes
- High resolution means
  - Fantastic new science



  - Difficulty detecting sources





- Sources are dispersed amongst many voxels
  - Voxel S/N is much less than integrated S/N
  - $S/N_{voxel} = 2\sim5 \times S/N_{source} / \sqrt{\#\ voxels}$
  - For instance, $49 < \#voxels < 1600$ results in $0.05\sim0.13 < S/N_{voxel} / S/N_{source} < 0.3\sim0.7$

Motivation

- For bright, compact sources
  - Individual voxels have okay/good S/N
  - Intensity threshold method can (and does!) work
- For extended or faint sources
  - Individual voxels have poor S/N
  - Intensity threshold method unlikely to work
  - Can be spatially unresolved but extended in frequency
- Intensity thresholding (Duchamp)
  - Will be incomplete
  - Is biassed
  - Refined reliability suffers trying to improve completeness
- WALLABY needs a source finder that complements Duchamp
  - Needs to work on groups of voxels
  - Should take advantage of high frequency resolution

CSIRO

New source finding methods

Novel source finding: General method

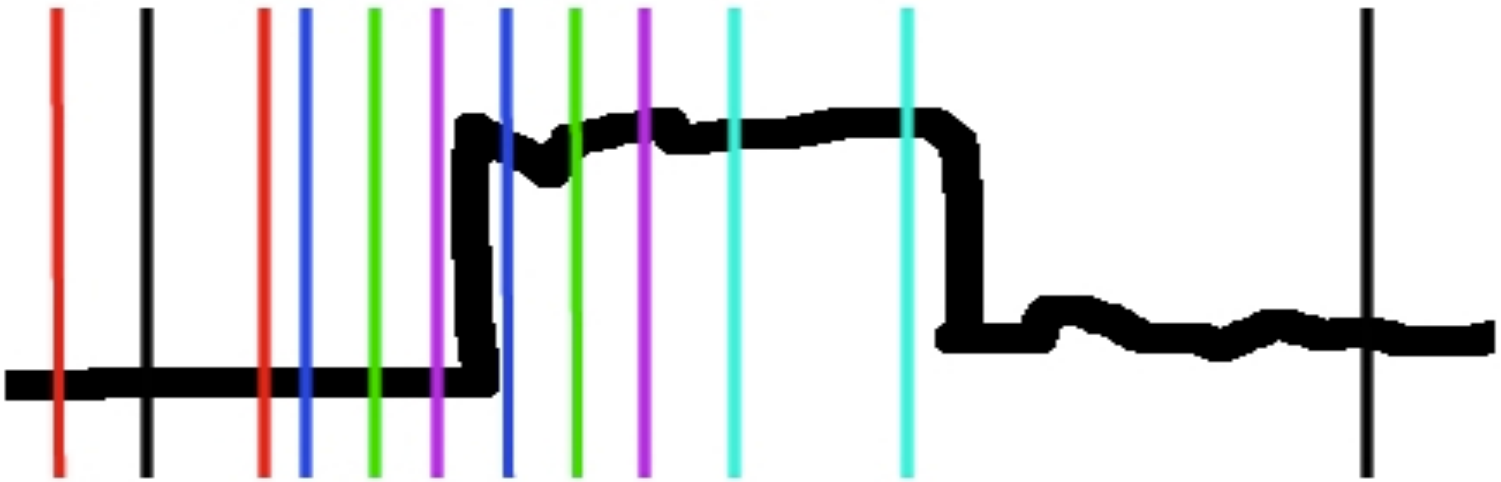
- Treating HI data cube as set of spectra
- Pre-process cube with 3-D iterative median smoothing.
- For each line of sight
  - Combine the neighbouring lines of sight
  - Use Wavelet analysis and Fourier transform to flatten and remove noise
  - Identify shapes in processed spectrum using one of:
    - Characterised noise method (under investigation)
    - Wavelet analysis (under investigation)
    - Gamma test
    - Intensity threshold
- Merging/Growing
- Source characterisation
- Source filtering

CSIRO

Iter. 3-D median smoothing: Purpose

- Goal: Reduce noise while preserving shape
  - Particularly edges
- Median works because it's non-linear
- Consider this,

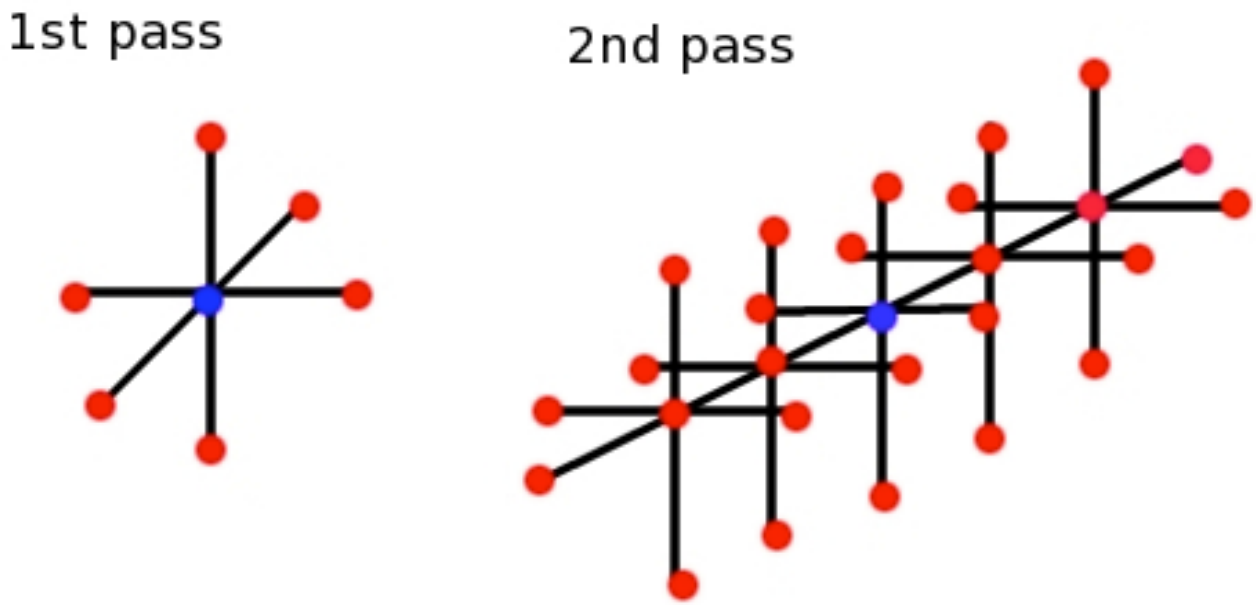
Iter. 3-D median smoothing: Purpose



- Away from edge
  - Median is all signal or all noise
- Near an edge
  - Contribution of signal and noise is weighted by how many elements lie within filter
- Best to use filter covering odd number of elements
- Better to use a filter that's too small than too big

Iter. 3-D median smoothing: Implementation

- Use a 3-D kernel
  - Maximises voxels used for median
- 1st pass = immediate neighbours
- 2nd pass = sausage in frequency direction
  - Highest resolution dimension

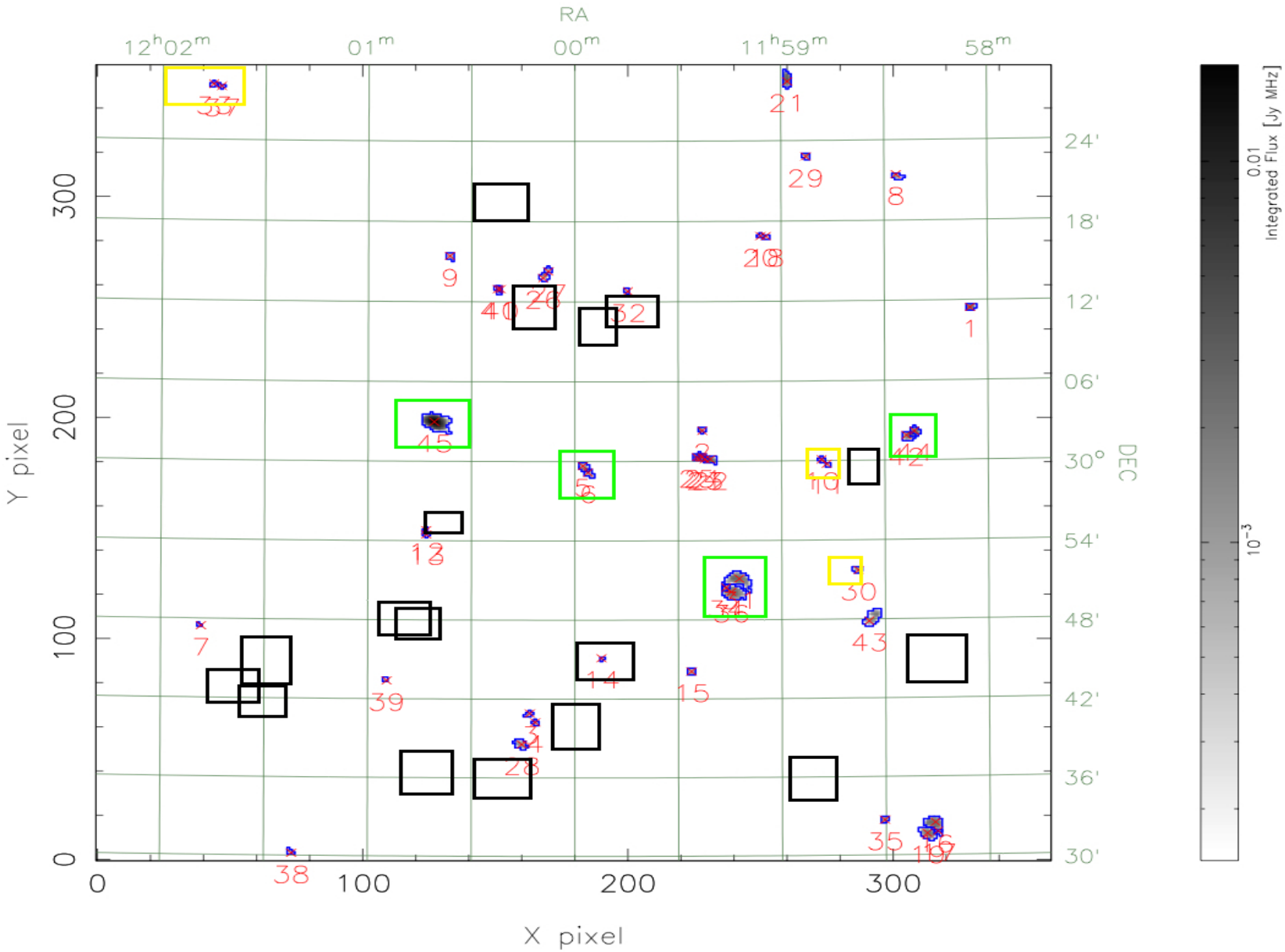


Iter. 3-D median smoothing: Results

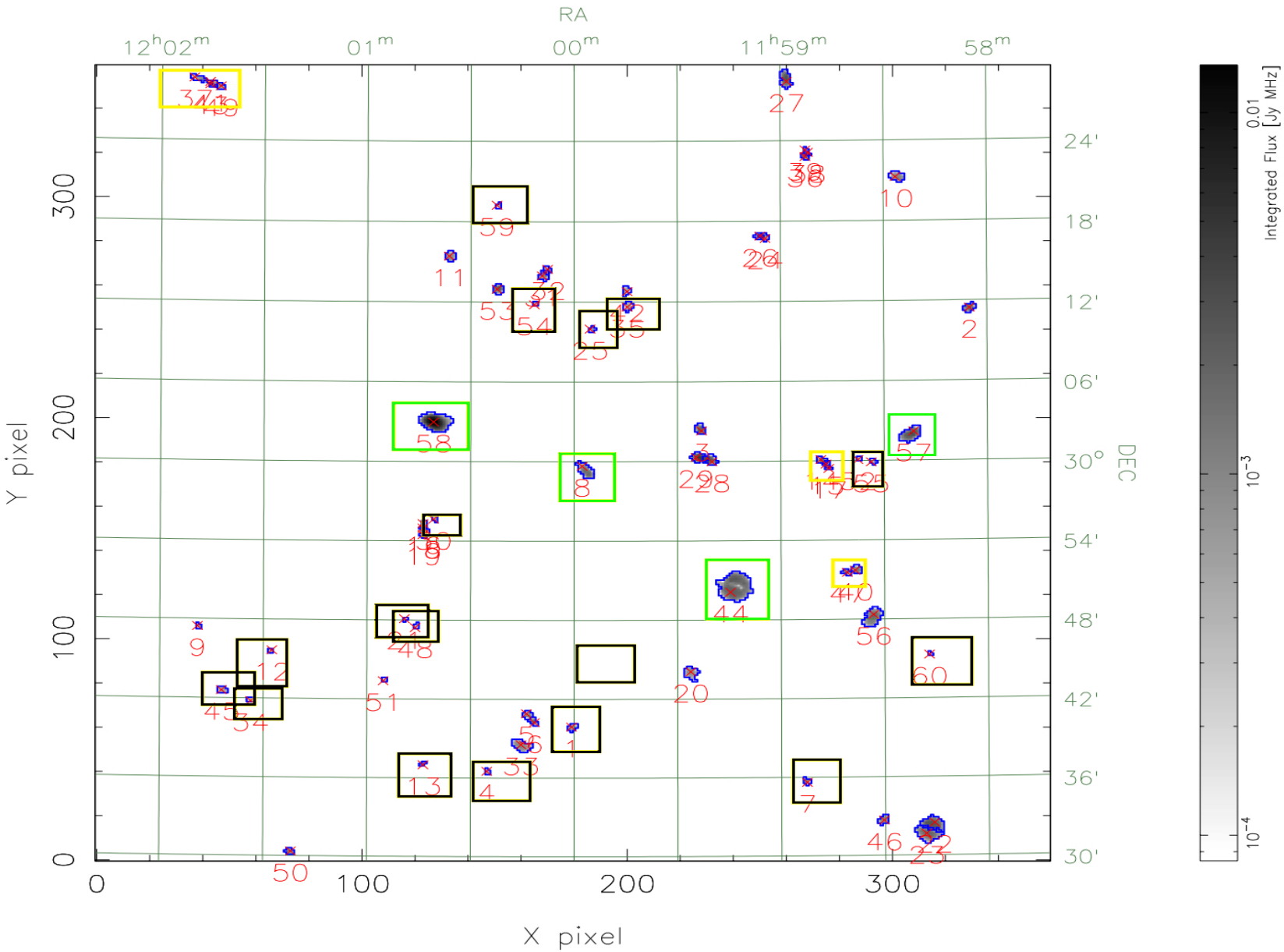
- Using Duchamp on Paolo Serra's Westerbork test cube
  - Find 50% more sources
  - Less fragmentation of sources
  - Slightly better/comparable to Duchamp's frequency smoothing/ wavelet analysis
- Confirmed that smoothing scale can't be too large
- Combination of smoothing and A'Trous wavelet reconstruction is fantastic!
  - Improved reliability, comparable otherwise.

CSIRO

Iter. 3-D median smoothing: Results



Iter. 3-D median smoothing: Results



CSIRO

Characterised noise method

- Datacubes are almost empty
  - Robust statistics `almost' solely characterise noise
  - Robust statistics include the median and the cumulative frequency distribution (c.f.d.)
- For a given line-of-sight
  - Noise is described by intensity c.f.d.
  - A group of pixels will have a statistically significantly different c.f.d. if they contain signal – searching for non-noise
- Implementation
  - Sliding box-car on different scales
  - Kuiper or Kolmogorov-Smirnov test
- Advantages
  - Ideal for extended sources
  - Estimates frequency extent
- Limitation
  - Correlated noise is selected as a source

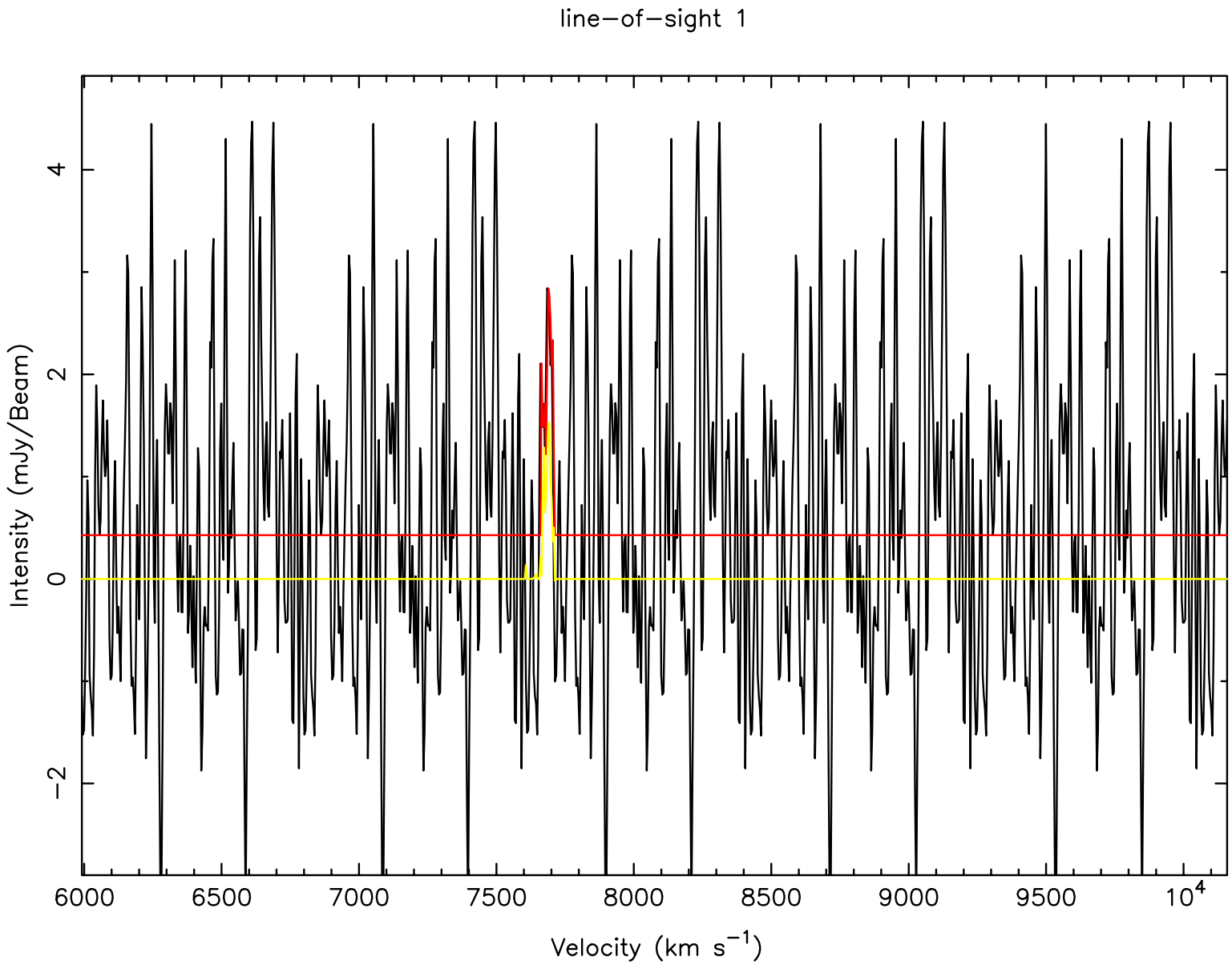
CSIRO

Characterised noise method: examples

- 12 lines-of-sight through the WSRT cube, where sources are known to exist
  - Good reliability and completeness
  - Does a good job of recovering frequency extent
  - Recovers sources amongst noise

CSIRO

Characterised noise method: examples

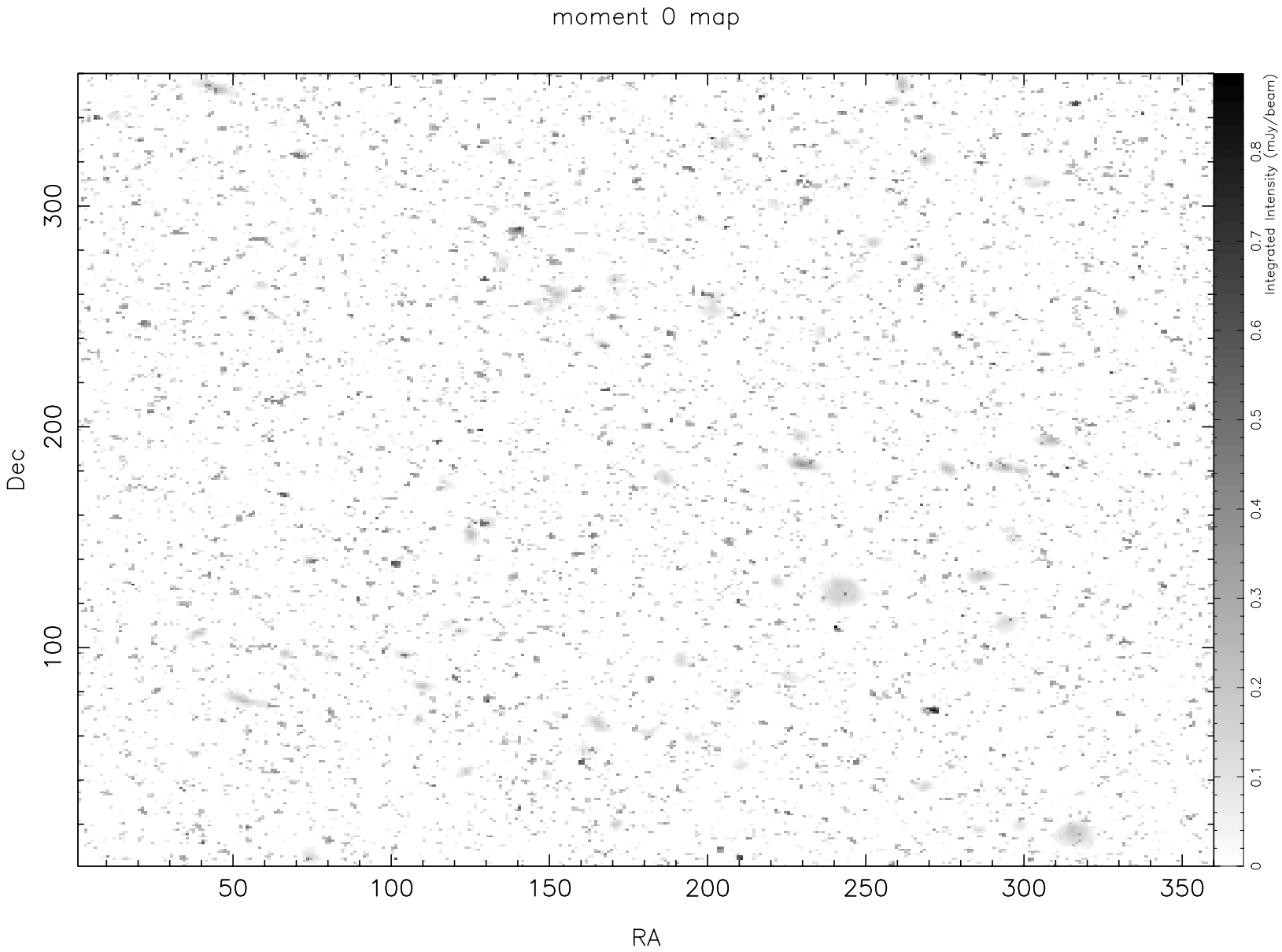


Characterised noise method: examples

- Blind search for sources in the WSRT cube
  - Recovers pretty much all sources
  - Recovers spatial extent
  - No preprocessing applied
  - No line-of-sight stacking used
  - No merging applied
  - No filtering applied
  - Poor implementation

CSIRO

Characterised noise method: examples



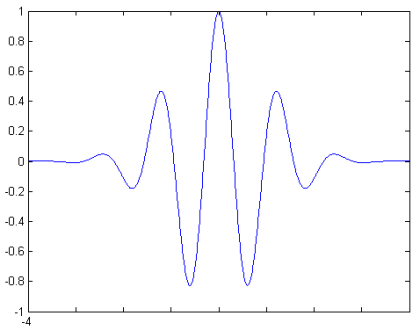
Wavelet method

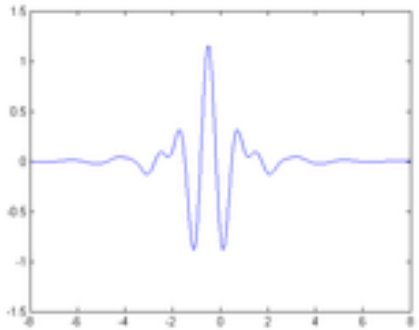
- Compute discrete wavelet transform for multiple basis functions
- Select significant wavelet coefficients
  - Absolute value cut
  - Relative cut using standard deviation threshold
- Determine best fitting basis out of significant coefficients
- Reconstruct signal using wavelet basis functions and coefficients
- Searching for shapes

CSIRO

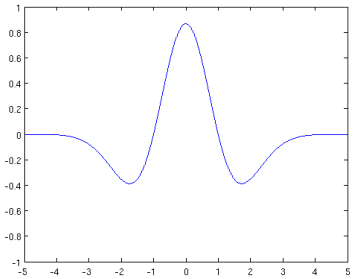
Wavelet method: wavelet library

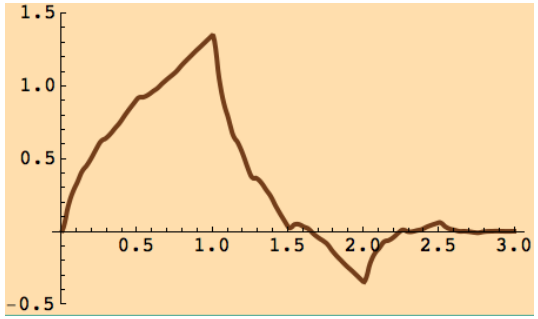
- Identified the following wavelets to try
  - Removing noise
    - Morlet
    - Meyer

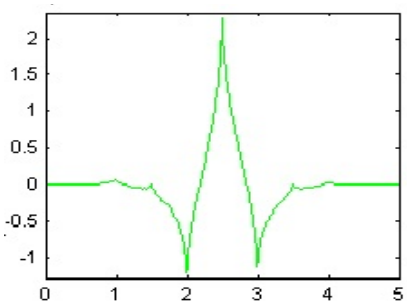


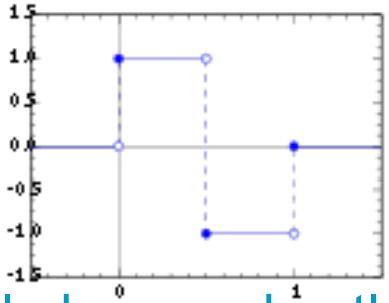


  - Finding sources
    - Mexican hat
    - Debauchies
    + mirrored
    - Coiflet
    - Haar









- Debauchies (+ mirrored) should identify double horns robustly

Wavelet method: Examples

- 12 lines-of-sight through the WSRT cube, where sources are known to exist
  - Ability to detect sources is critically dependent upon choice of basis
  - Only detects sources matching basis function
  - Large scales are useless
  - Small scales can be filtered out
  - Blindingly fast!
- Experimenting with best approach for identifying source components
- Still working on best way to make use of multiple basis functions

CSIRO

Summary

- 3-D iterative median smoothing works and has benefits for source finding/characterisation
- Successful proof-of-concept for
  - Characterised noise method
  - Wavelet method
- Future work
  - Improved implementations
  - Tweaks to both methods
  - Investigate integration of both methods + intensity thresholding

CSIRO