Source Extraction and Characterisation II

Spectral Line Emission

Tobias Westmeier (ICRAR / UWA)

ICRAR
International Centre for Radio Astronomy Research

Curtin University

THE UNIVERSITY OF WESTERN AUSTRALIA

Government of Western Australia
Department of the Premier and Cabinet
Office of Science

Outline of Lecture

- ★ Source finding
  - 3D source finding
  - Software
  - Metrics
  - Algorithms
- ★ Source parameterisation
  - Basic parameters
  - Moment analysis
  - Spectral fitting
  - Frequency – redshift – velocity
  - Uncertainties

Source Finding

Source Finding – 3D Approach

★ **Assumptions**

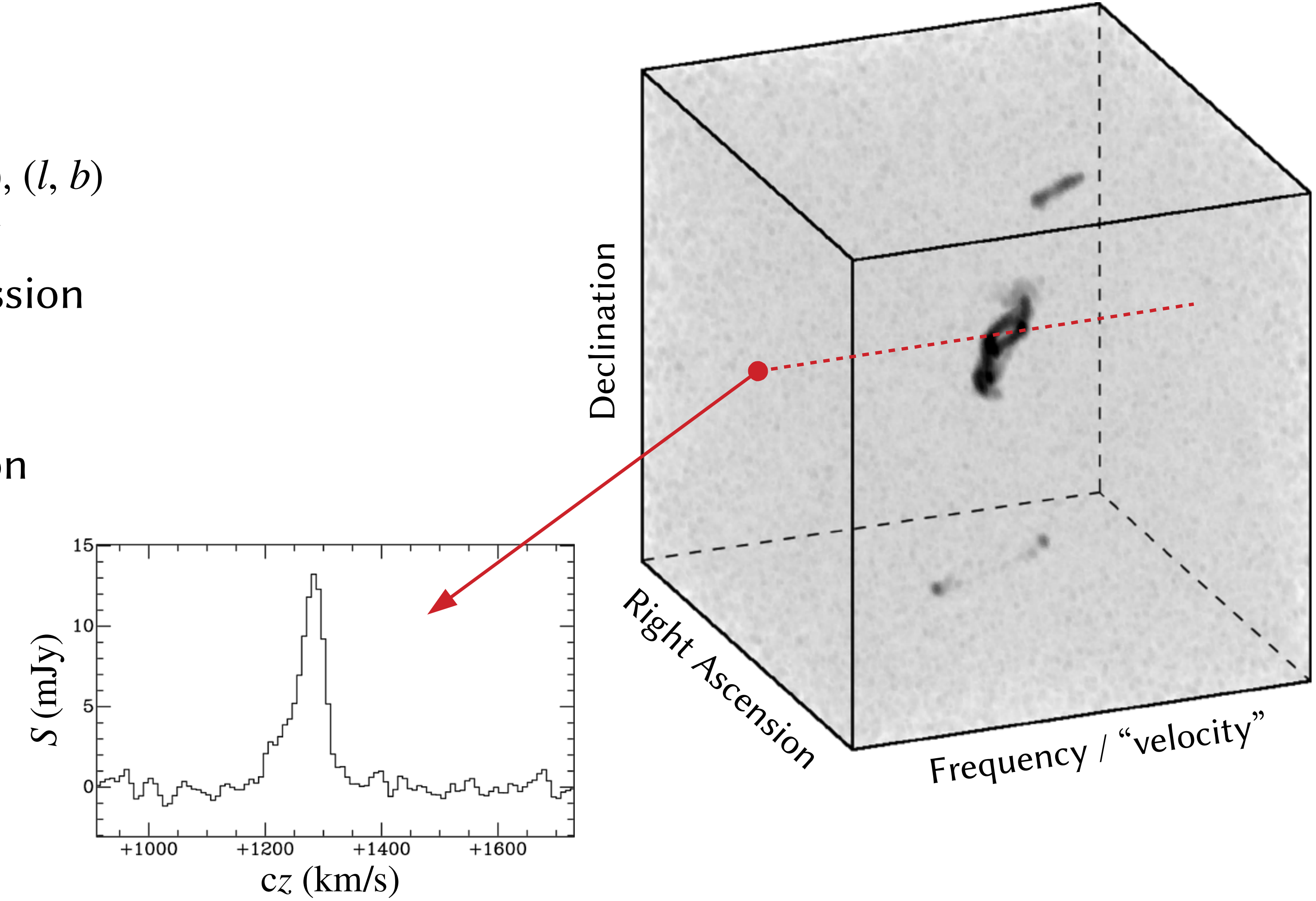
- 3D image data cubes
  - 2 spatial dimensions: (α, δ), (l, b)
  - 1 spectral dimension: f, v, z
- Gaussian noise + source emission

★ **Advantages**

- Redshift / distance information
- Less source confusion

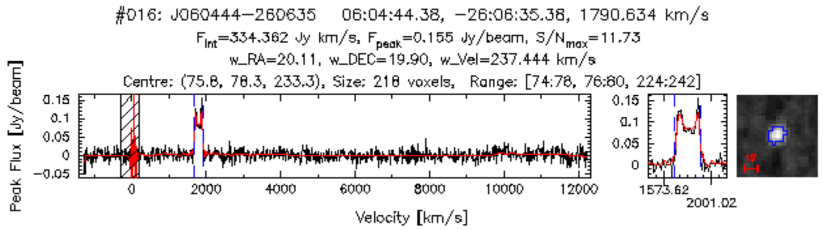
★ **Disadvantages**

- Larger data volume
- 3D approach required



★ Software

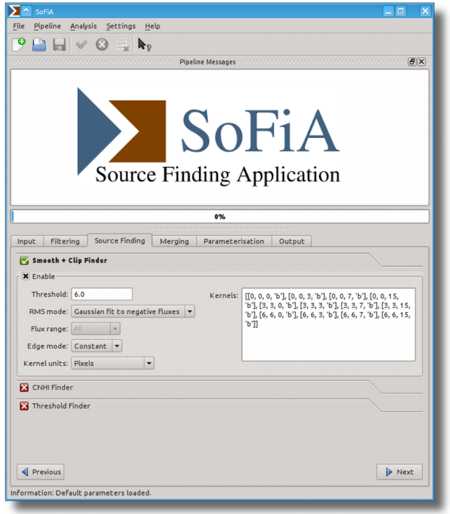
▶ Duchamp / Sélavy



- 3D source finder implemented in the ASKAPsoft pipeline
- Developed by Matthew Whiting
- Duchamp: https://www.atnf.csiro.au/people/Matthew.Whiting/Duchamp/
- Sélavy: https://www.atnf.csiro.au/computing/software/askapsoft/sdp/docs/current/analysis/

▶ SoFiA (*Source Finding Application*)

- Stand-alone 3D source finding pipeline
- Originally developed for extragalactic H I surveys
- Graphical user interface
- GitHub: https://github.com/SoFiA-Admin/SoFiA/
- SoFiA wiki: https://github.com/SoFiA-Admin/SoFiA/wiki



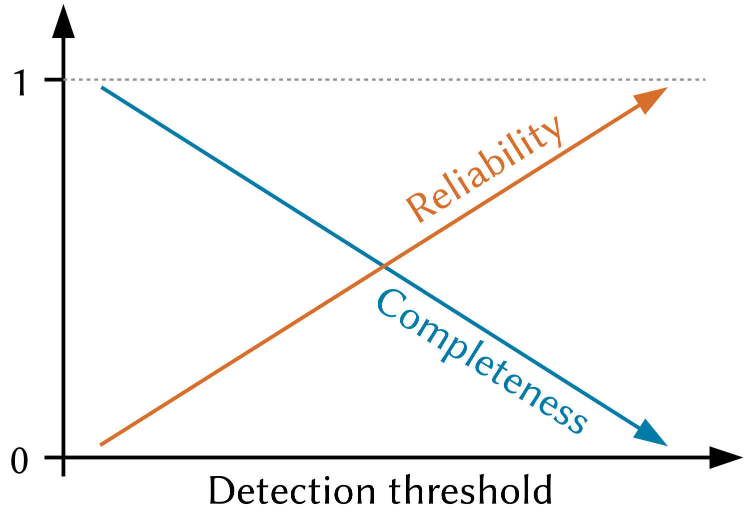
Source Finding – Metrics

★ Source finding

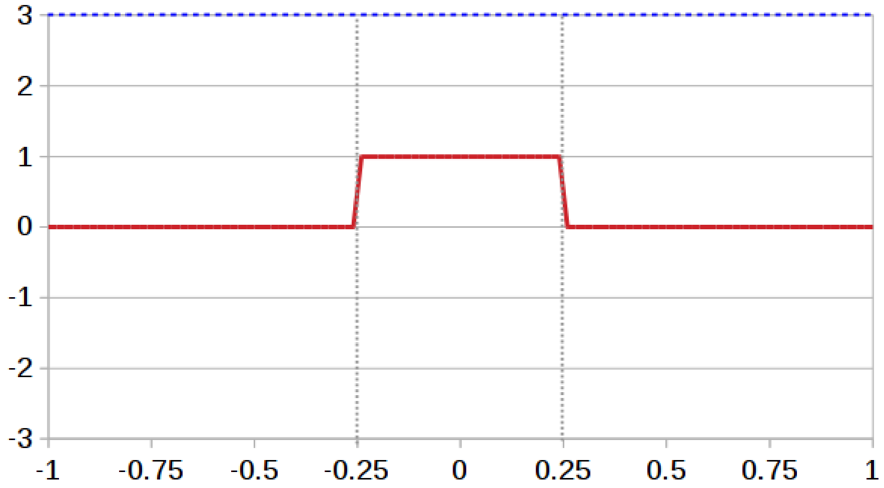
- Detection of signal in data containing statistical noise
- WALLABY: 500,000 galaxies, 1 PB of data $\longrightarrow$ $\boxed{\text{automation required}}$

★ Metrics

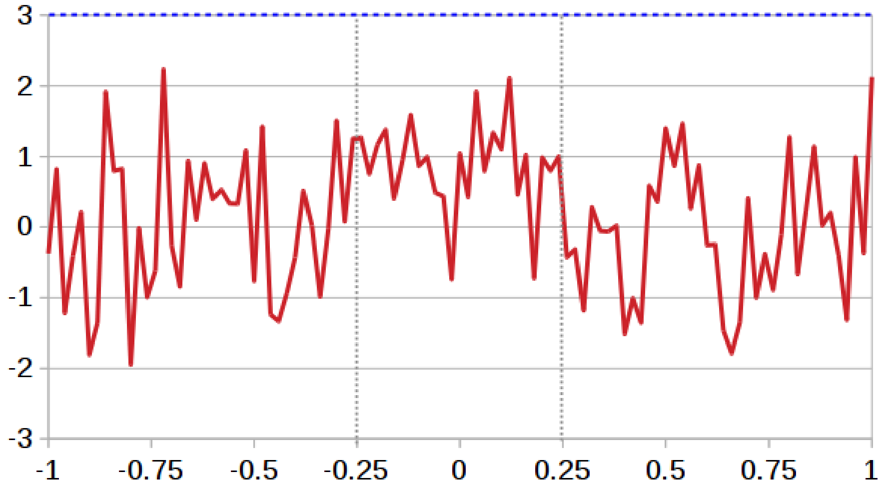
- Completeness
  - Fraction of sources detected $\longrightarrow$ $C = \text{True} / \text{All}$
- Reliability
  - Fraction of genuine detections $\longrightarrow$ $R = \text{True} / (\text{False} + \text{True})$
- Function of signal-to-noise ratio
- Compromise between
  - Low threshold $\longrightarrow$ high completeness, but false detections
  - High threshold $\longrightarrow$ high reliability, but missing sources



- Signal-to-noise ratio (SNR)
  - Simple 1D example
    - Box-shaped source of $S = 1$, $w = 25$

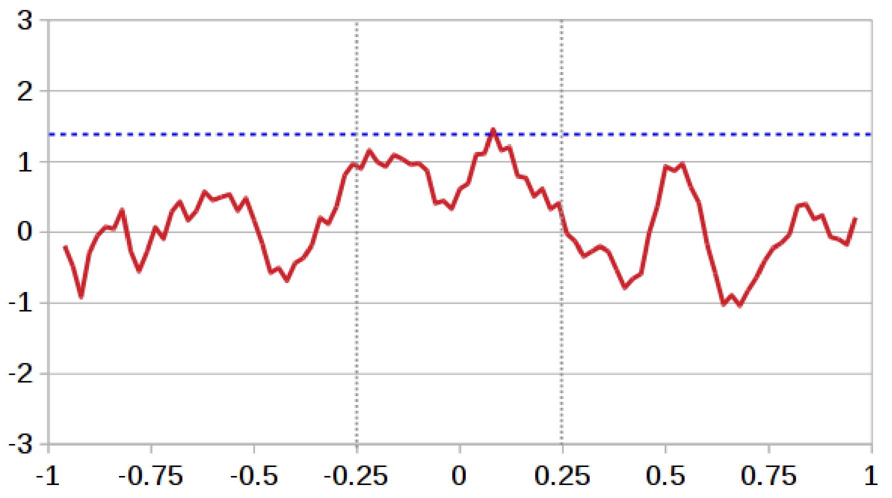


- ★ Signal-to-noise ratio (SNR)
  - ▸ Simple 1D example
    - Box-shaped source of $S = 1$, $w = 25$
    - Add Gaussian noise of $\sigma = 1$
  - ▸ Convolve with boxcar filter
    - Original $\longrightarrow$ $\sigma = 1.00$, SNR $= 1.00$ ($\mathrm{SNR}_{\mathrm{int}} = 5.00$)

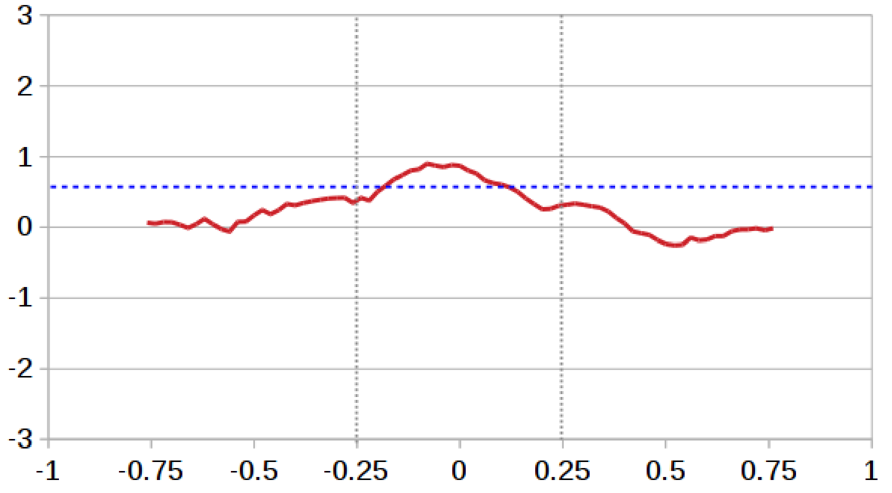


Source Finding – Smoothing and Clipping

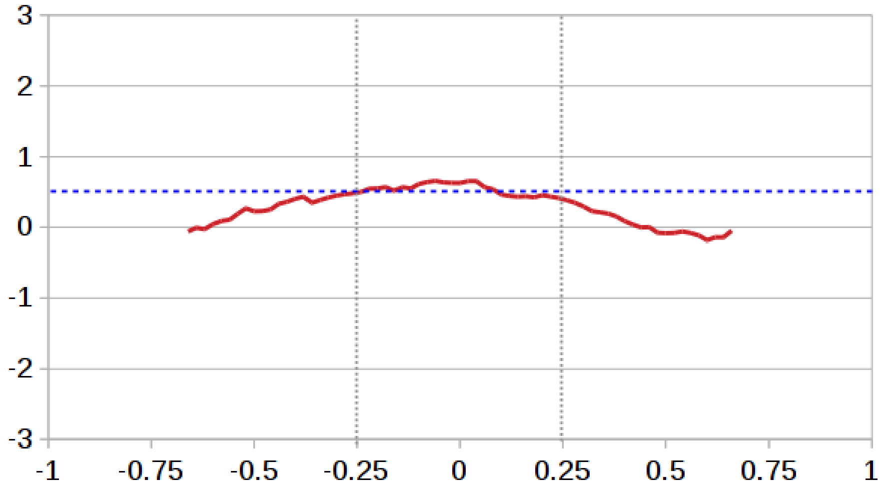
- ★ Signal-to-noise ratio (SNR)
  - ▸ Simple 1D example
    - Box-shaped source of $S = 1$, $w = 25$
    - Add Gaussian noise of $\sigma = 1$
  - ▸ Convolve with boxcar filter
    - Original $\longrightarrow$ $\sigma = 1.00$, SNR $= 1.00$ ($\mathrm{SNR}_{\mathrm{int}} = 5.00$)
    - Size 5 $\longrightarrow$ $\sigma = 0.45$, SNR $= 2.24$



- ★ Signal-to-noise ratio (SNR)
  - Simple 1D example
    - Box-shaped source of $S = 1$, $w = 25$
    - Add Gaussian noise of $\sigma = 1$
  - Convolve with boxcar filter
    - Original $\longrightarrow$ $\sigma = 1.00$, SNR = 1.00 (SNR$_{\mathrm{int}}$ = 5.00)
    - Size 5 $\longrightarrow$ $\sigma = 0.45$, SNR = 2.24
    - Size 25 $\longrightarrow$ $\sigma = 0.20$, SNR = 5.00



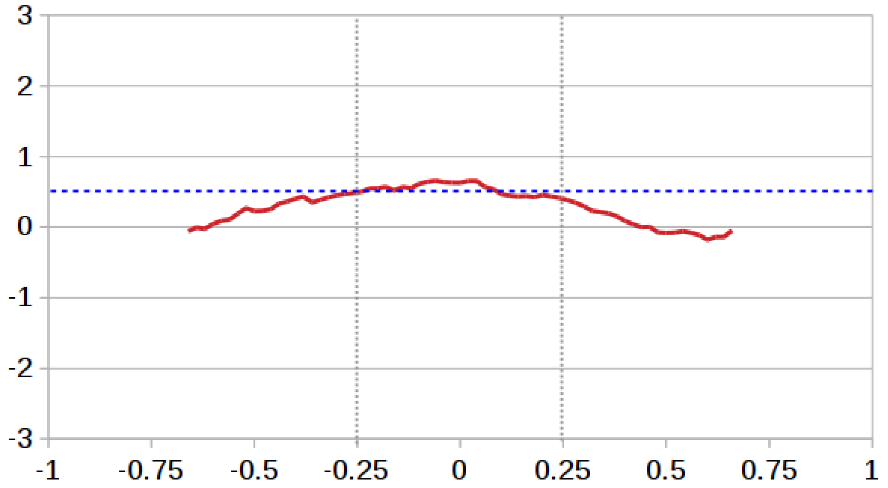
- ★ Signal-to-noise ratio (SNR)
  - Simple 1D example
    - Box-shaped source of $S = 1$, $w = 25$
    - Add Gaussian noise of $\sigma = 1$
  - Convolve with boxcar filter
    - Original $\longrightarrow$ $\sigma = 1.00$, SNR = 1.00 ($\mathrm{SNR}_{\mathrm{int}} = 5.00$)
    - Size 5 $\longrightarrow$ $\sigma = 0.45$, SNR = 2.24
    - Size 25 $\longrightarrow$ $\sigma = 0.20$, SNR = 5.00
    - Size 35 $\longrightarrow$ $\sigma = 0.17$, SNR = 4.23

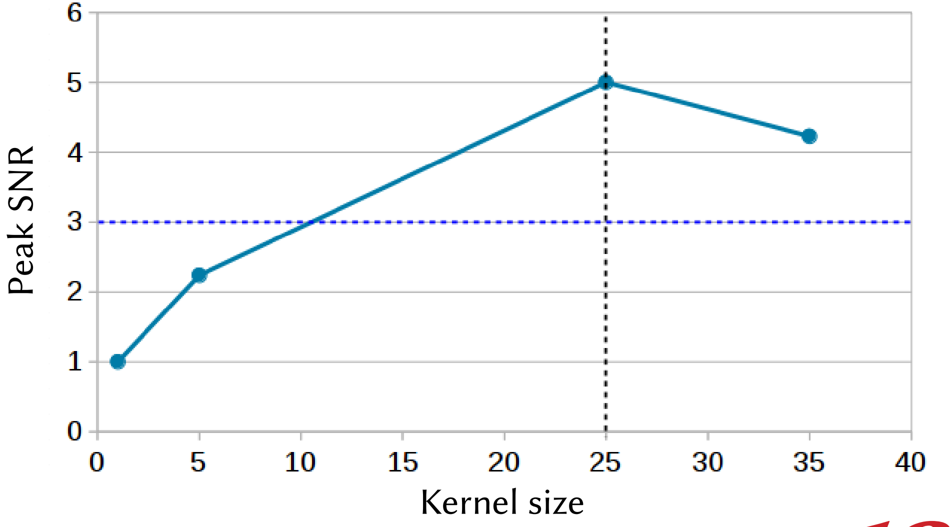


Source Finding – Smoothing and Clipping

★ Signal-to-noise ratio (SNR)

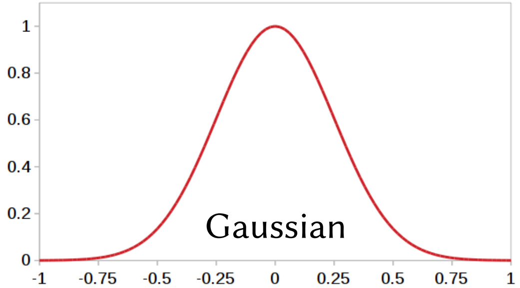
- Simple 1D example
  - Box-shaped source of $S = 1$, $w = 25$
  - Add Gaussian noise of $\sigma = 1$
- Convolve with boxcar filter
  - Original $\longrightarrow$ $\sigma = 1.00$, SNR = 1.00 ($\mathrm{SNR}_{\mathrm{int}} = 5.00$)
  - Size 5 $\longrightarrow$ $\sigma = 0.45$, SNR = 2.24
  - Size 25 $\longrightarrow$ $\sigma = 0.20$, SNR = 5.00
  - Size 35 $\longrightarrow$ $\sigma = 0.17$, SNR = 4.23
- Conclusions
  - Smooth data to optimal resolution to maximise SNR of sources
  - Recovery of integrated SNR (± noise) for kernels that match shape and size of source

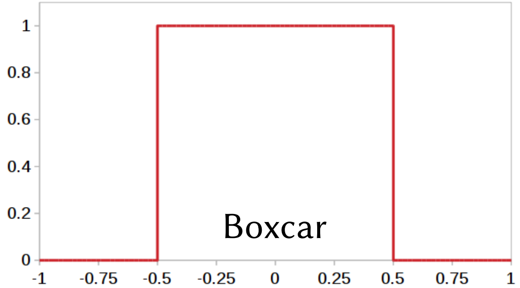


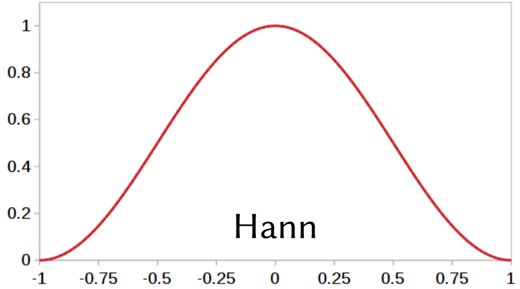


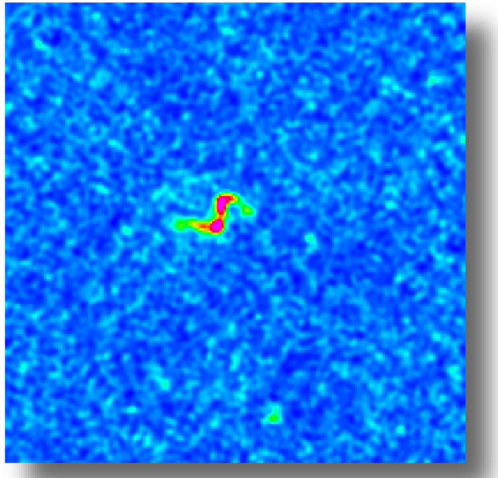
Source Finding – Smoothing and Clipping

- Smooth + clip algorithm
  - Convolution with multiple 3D kernels for spatial and spectral smoothing on different scales
  - Measure RMS on each scale and apply threshold of $N \times \mathrm{RMS}$
  - Add pixels above threshold to source mask

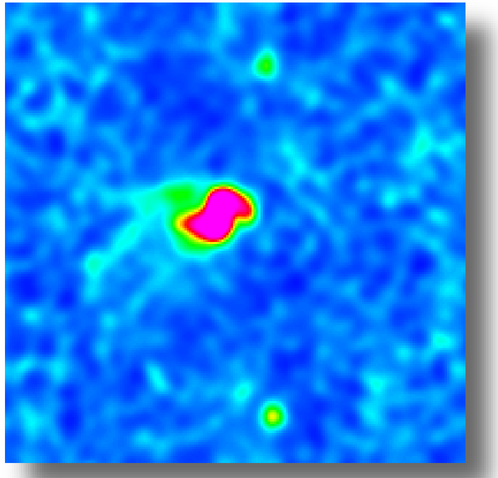




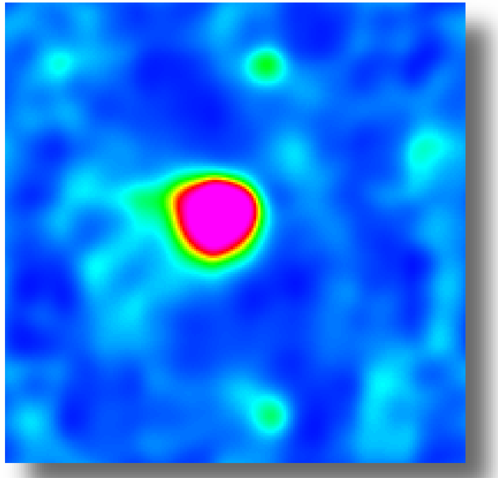




No smoothing



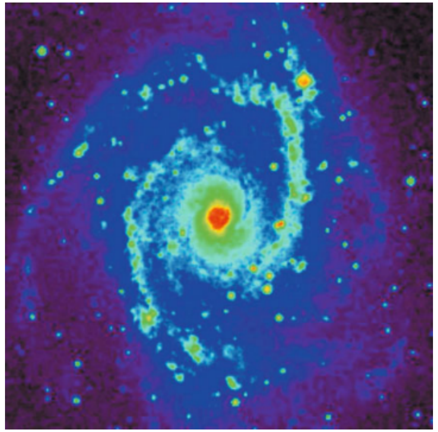
Gaussian of σ = 3 pixels

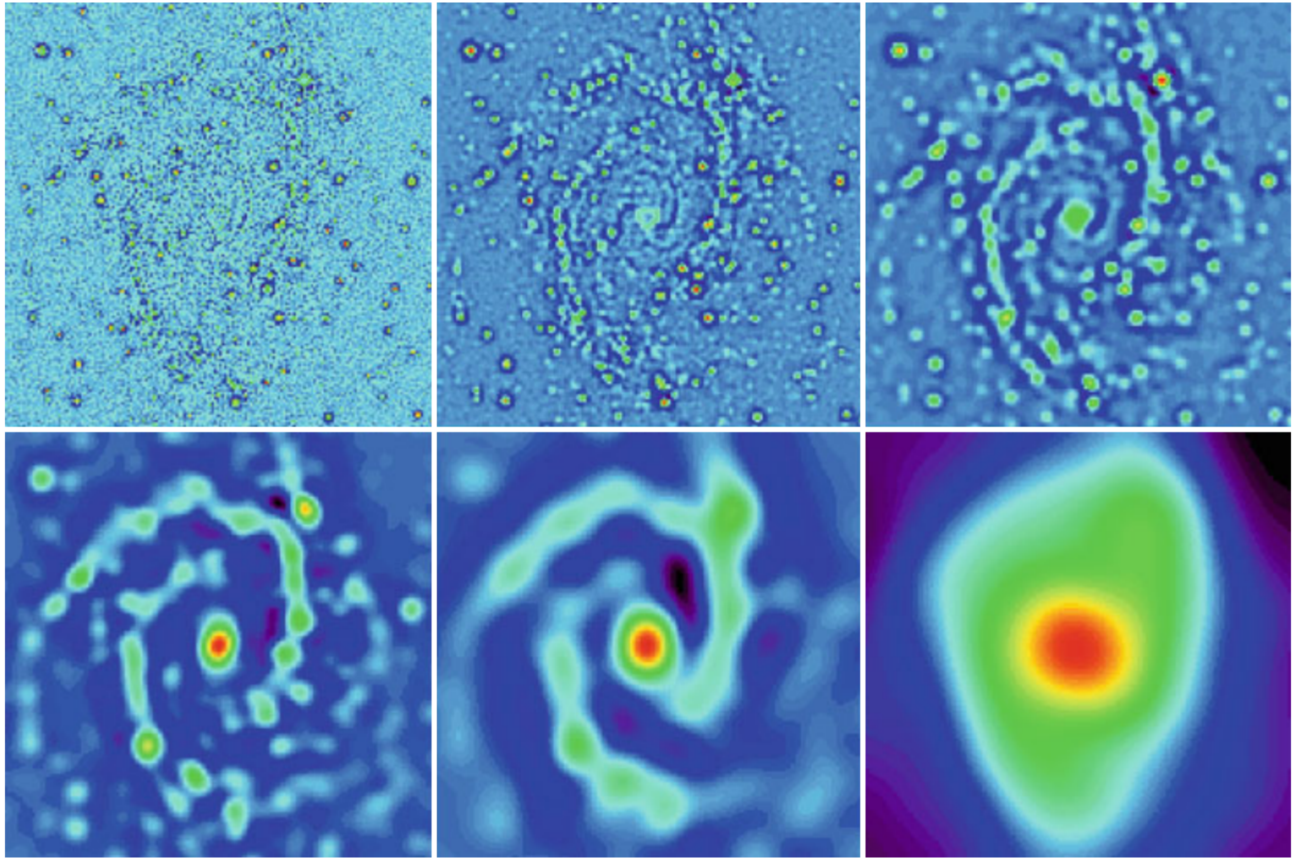


Gaussian of σ = 6 pixels

- Alternative algorithms
  - Wavelet decomposition

$$D(x) = c_J(x) + \sum_{j=1}^{J} w_j(x)$$


NGC 2997


NGC 2997 – wavelet transform

Starck, Murtagh & Bertero (2011)

Source Finding – Wavelet Decomposition

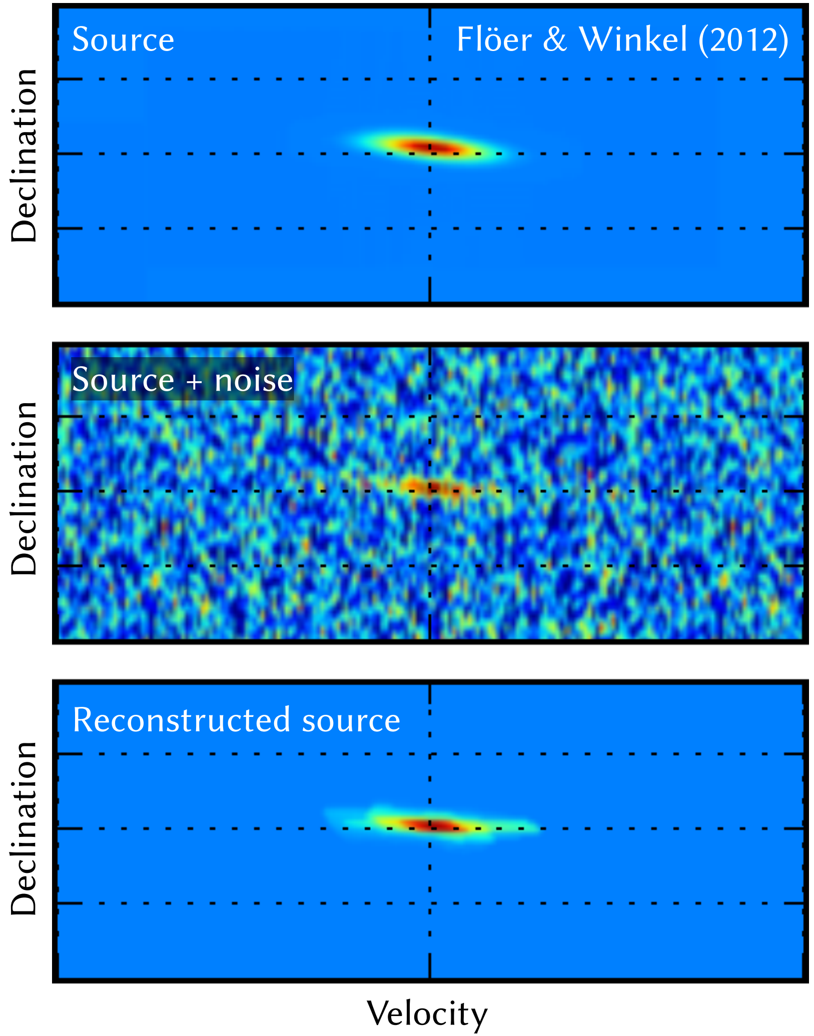
- ★ Alternative algorithms
  - Wavelet decomposition

$$D(x) = c_J(x) + \sum_{j=1}^{J} w_j(x)$$

  - 2D–1D wavelet decomposition

$$D(x) = c_{J_1,J_2}(x) + \sum_{j_1} w_{j_1,J_2}(x) + \sum_{j_2} w_{J_1,j_2}(x) + \sum_{j_1,j_2} w_{j_1,j_2}(x)$$

  - See Flöer & Winkel (2012) for details



★ Alternative algorithms

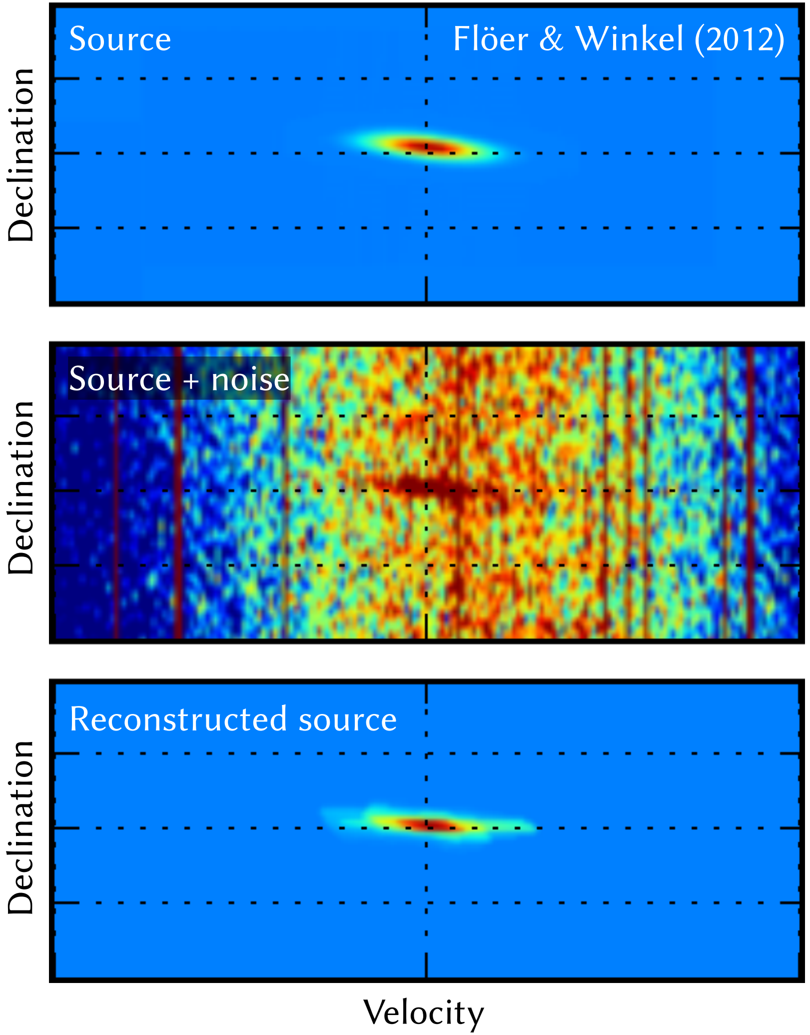
- Wavelet decomposition

$$D(x) = c_J(x) + \sum_{j=1}^{J} w_j(x)$$

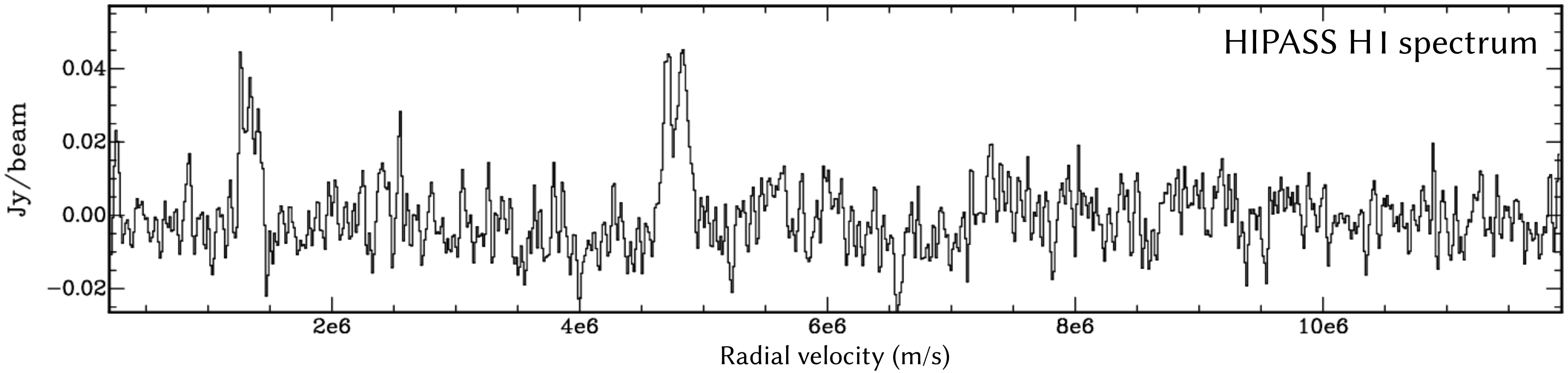
- 2D–1D wavelet decomposition

$$\begin{aligned} D(x) &= c_{J_1,J_2}(x) \\ &\quad + \sum_{j_1} w_{j_1,J_2}(x) + \sum_{j_2} w_{J_1,j_2}(x) \\ &\quad + \sum_{j_1,j_2} w_{j_1,j_2}(x) \end{aligned}$$

- See Flöer & Winkel (2012) for details



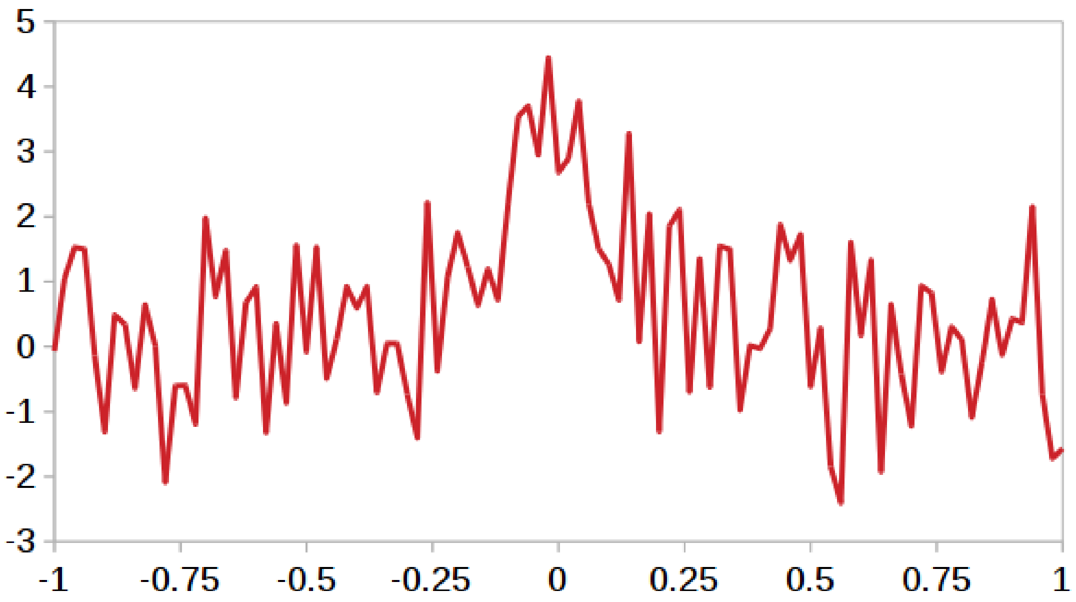
- Alternative algorithms
  - Sources may be spatially unresolved
    - Source finding problem reduces from 3D to 1D
  - Characterised Noise H I source finder (CNHI)
    - Kuiper test in a running window along the spectral axis
      $\longrightarrow$ uncover regions statistically inconsistent with pure Gaussian noise
  - See Jurek (2012) for more details



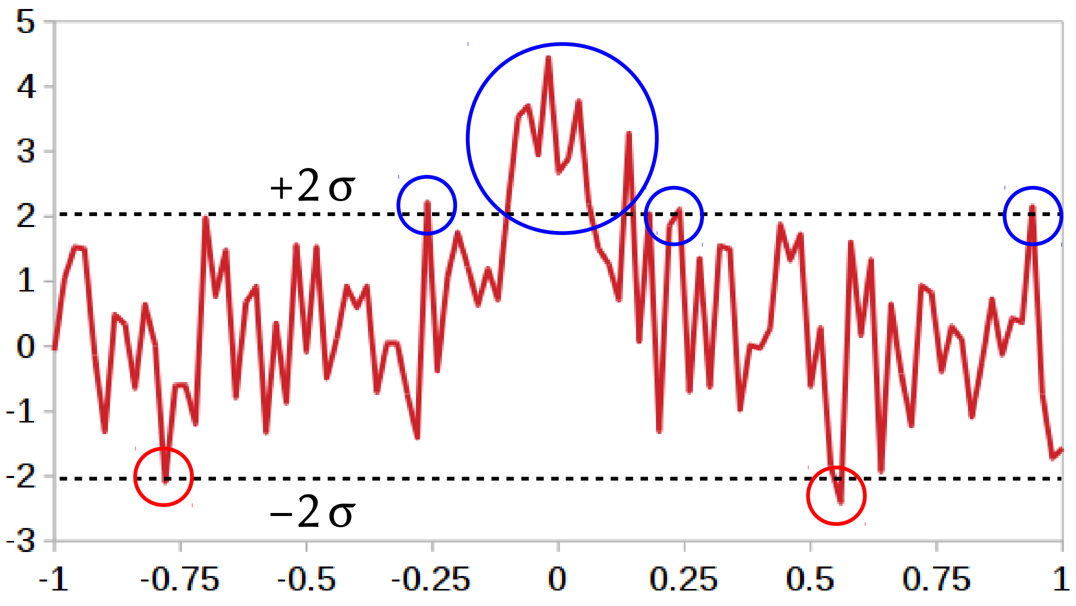
Source Finding – Reliability

★ Estimating reliability

- Fundamental assumptions
  - Gaussian noise, no offset
  - Astronomical signal is positive
    (e.g. H I emission)
  - No artefacts
    (e.g. RFI, sidelobes, continuum residuals)



- ★ Estimating reliability
  - ▸ Fundamental assumptions
    - Gaussian noise, no offset
    - Astronomical signal is positive (e.g. H I emission)
    - No artefacts (e.g. RFI, sidelobes, continuum residuals)
- ★ Method
  - ▸ Search for all signals with $|S| > N \times \sigma$
    - $S < 0 \quad \longrightarrow \quad$ *Noise*
    - $S > 0 \quad \longrightarrow \quad$ *Noise or source*

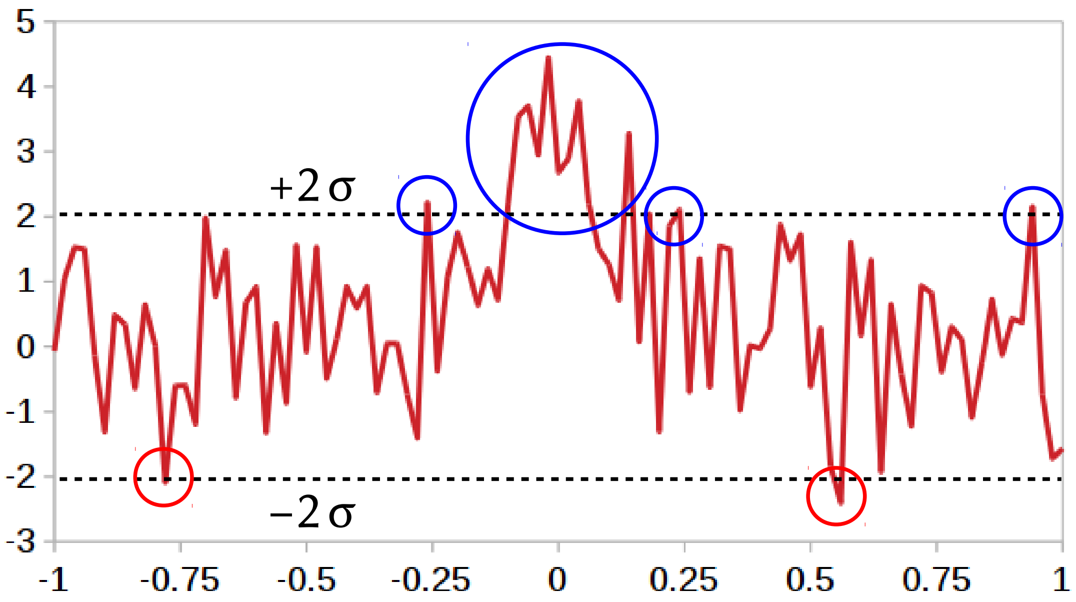


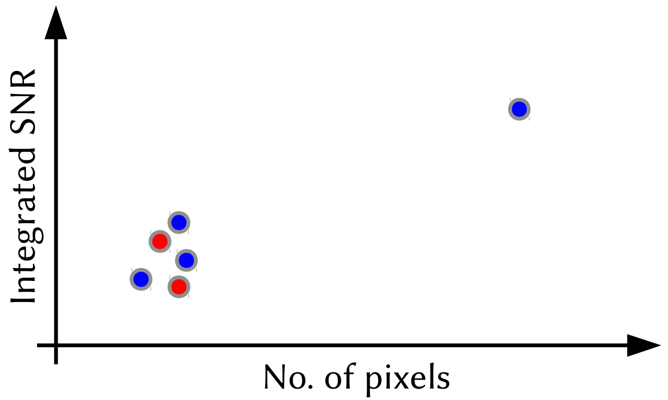
★ **Estimating reliability**

- Fundamental assumptions
  - Gaussian noise, no offset
  - Astronomical signal is positive
    (e.g. H I emission)
  - No artefacts
    (e.g. RFI, sidelobes, continuum residuals)

★ **Method**

- Search for all signals with $|S| > N \times \sigma$
  - $S < 0 \longrightarrow$ *Noise*
  - $S > 0 \longrightarrow$ *Noise or source*
- Compare density in parameter space



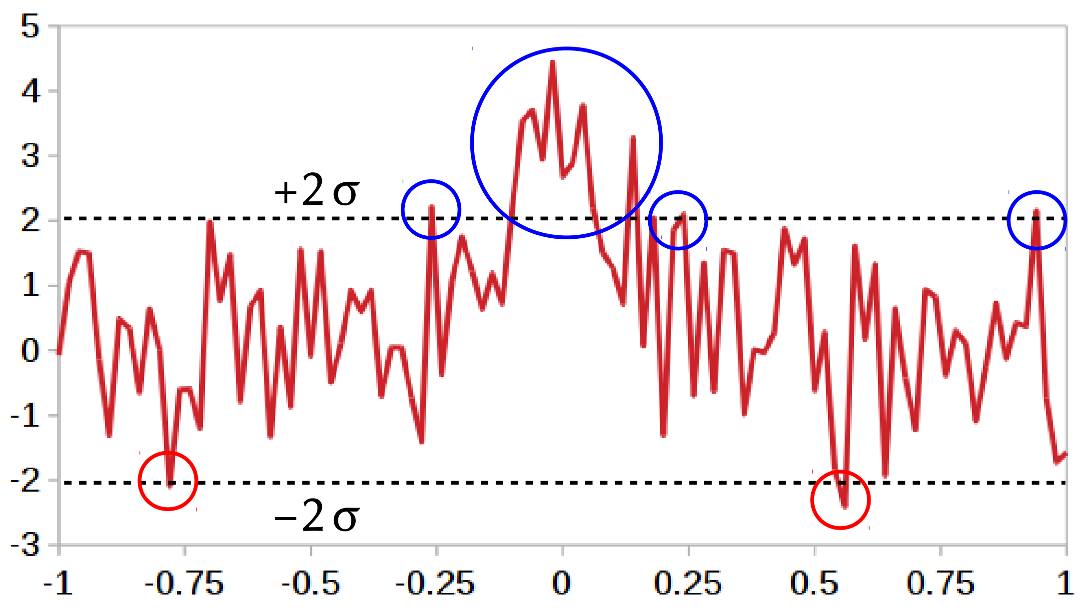


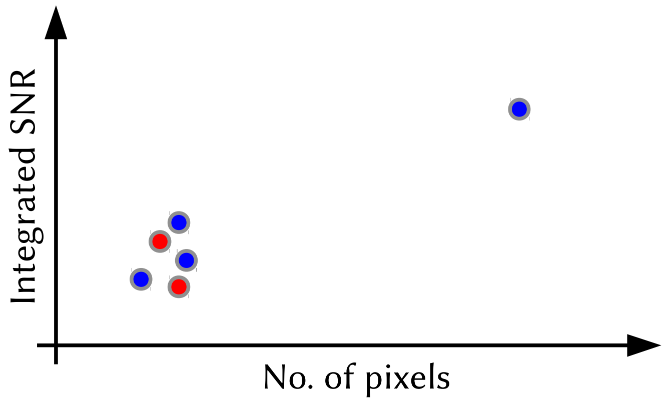
★ Estimating reliability

- Fundamental assumptions
  - Gaussian noise, no offset
  - Astronomical signal is positive (e.g. H I emission)
  - No artefacts (e.g. RFI, sidelobes, continuum residuals)

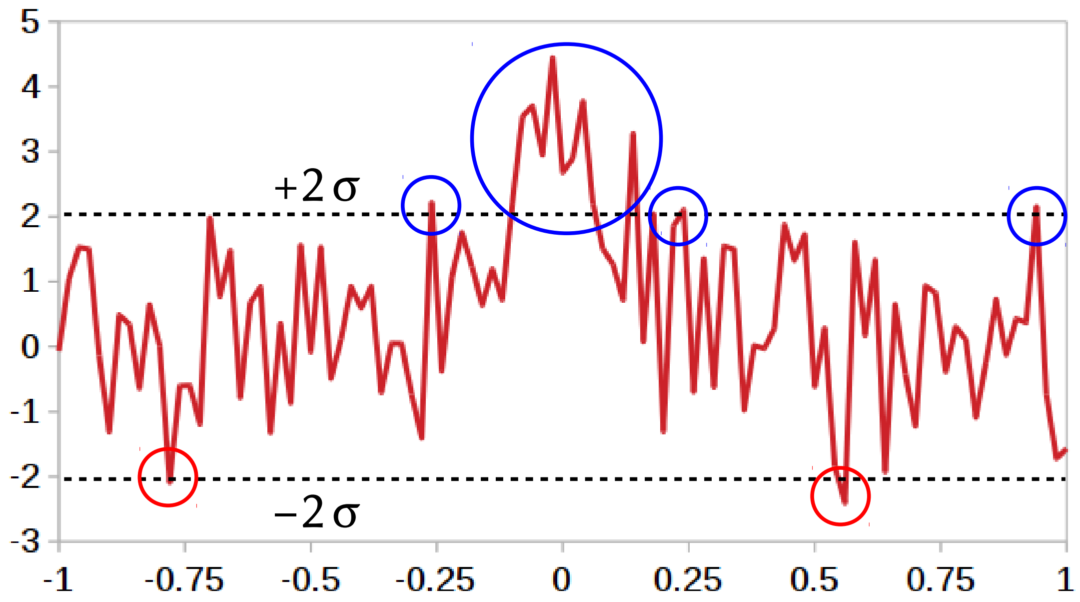
★ Method

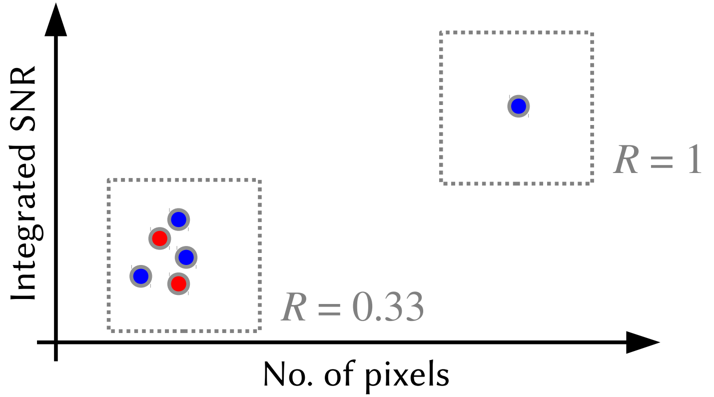
- Search for all signals with $|S| > N \times \sigma$
  - $S < 0 \;\; \longrightarrow \;\;$ *Noise*
  - $S > 0 \;\; \longrightarrow \;\;$ *Noise or source*
- Compare density in parameter space
  - Reliability $\boxed{R \equiv \dfrac{T}{T+F} \;\; \rightarrow \;\; \dfrac{P-N}{P}}$





- ★ **Estimating reliability**
  - ▸ Fundamental assumptions
    - Gaussian noise, no offset
    - Astronomical signal is positive (e.g. H I emission)
    - No artefacts (e.g. RFI, sidelobes, continuum residuals)
- ★ **Method**
  - ▸ Search for all signals with $|S| > N \times \sigma$
    - $S < 0 \quad \rightarrow \quad$ *Noise*
    - $S > 0 \quad \rightarrow \quad$ *Noise or source*
  - ▸ Compare density in parameter space
    - Reliability $\boxed{R \equiv \frac{T}{T+F} \rightarrow \frac{P-N}{P}}$

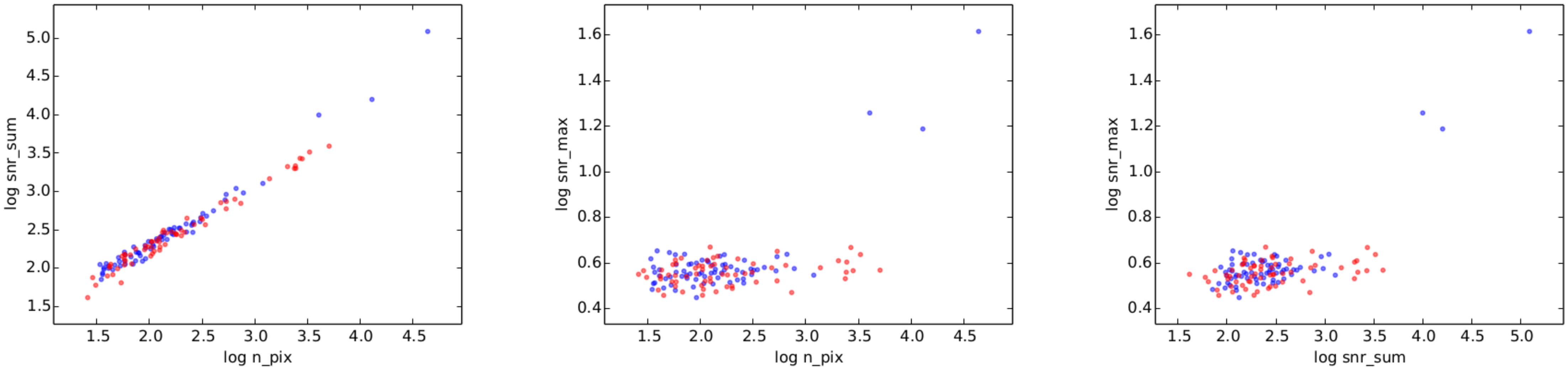




Estimating reliability

Serra et al. 2012, PASA, 29, 296

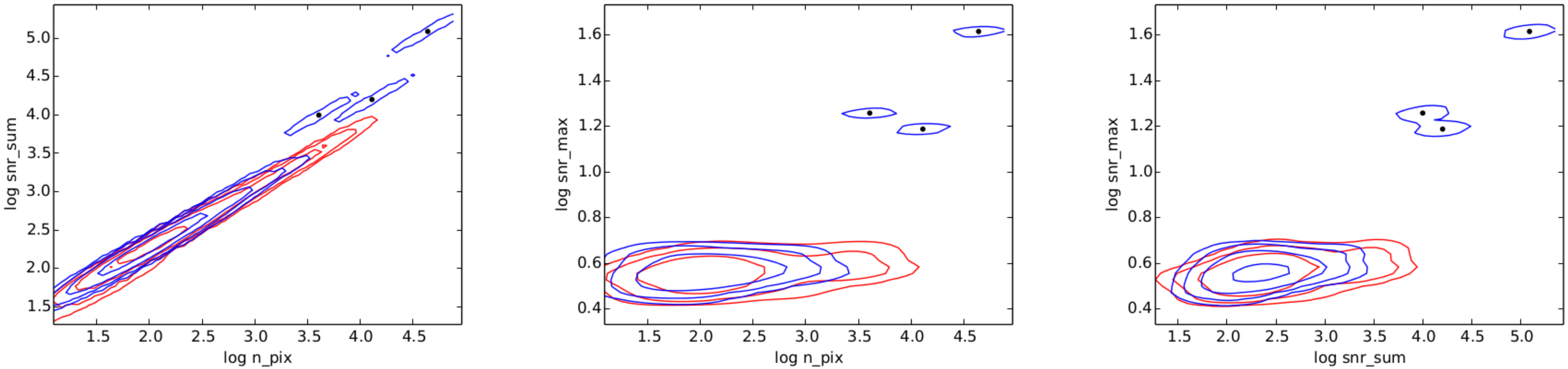
- Retain sources above a meaningful threshold, e.g. $R > 0.9$



★ Estimating reliability

Serra et al. 2012, PASA, 29, 296

- Retain sources above a meaningful threshold, e.g. $R > 0.9$

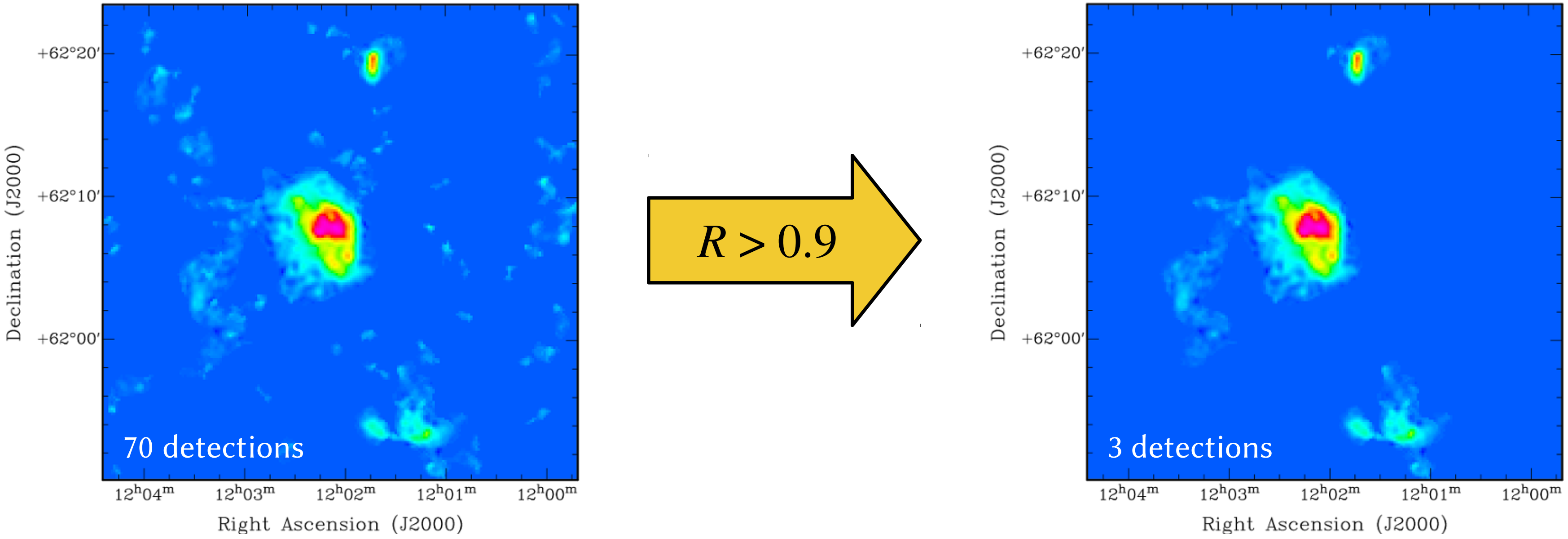


- Highly reliable source catalogue
  - Dozens of false detections removed as unreliable
  - Enables use of low source finding threshold of $\approx 3\,\sigma$

- ★ **Estimating reliability**

Serra et al. 2012, PASA, 29, 296

  - Retain sources above a meaningful threshold, e.g. $R > 0.9$



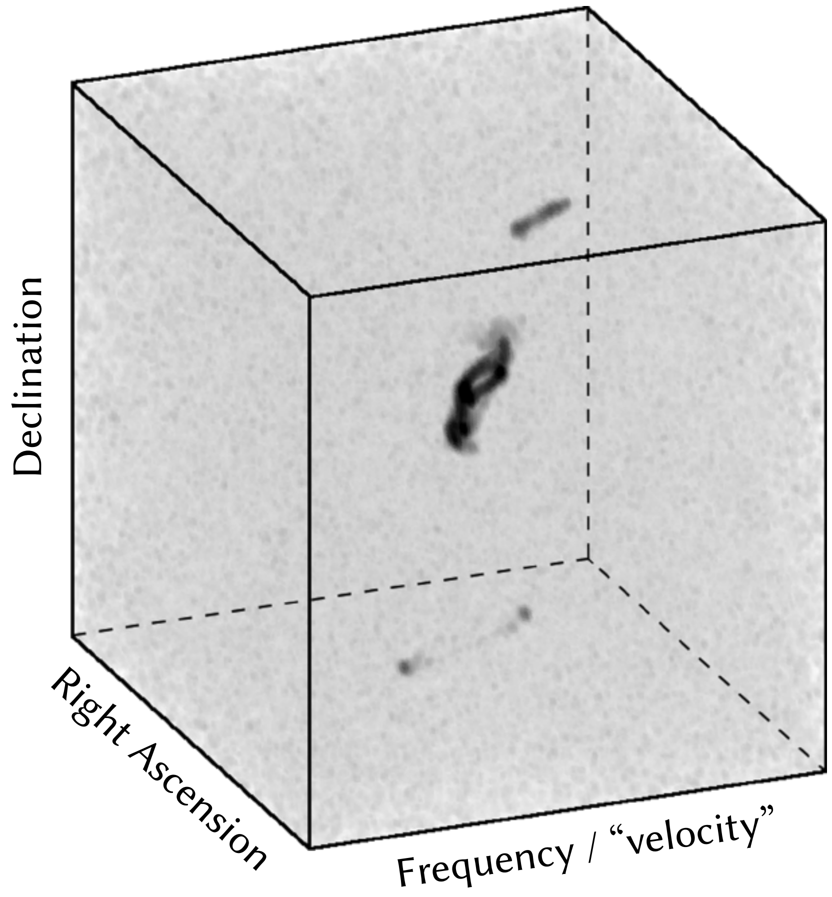
  - Highly reliable source catalogue
    - Dozens of false detections removed as unreliable
    - Enables use of low source finding threshold of $\approx 3\,\sigma$

Requires clean data with Gaussian noise plus source emission and no artefacts!

Parameterisation

★ Source parameterisation

- Process of measuring the basic observational parameters of a source
  - Position → *sky position, frequency / radial velocity*
  - Size → *angular size, spectral line width*
  - Flux → *peak flux density / brightness temperature, integrated flux*
  - Other → *orientation, morphology, asymmetry, etc.*
- Conversion to physical parameters
  - Position → *redshift, distance*
  - Size → *diameter, rotation velocity, temperature*
  - Flux → *luminosity, column density, mass*
- Effect of noise
  - Statistical uncertainty
  - Parameterisation often dominated by systematic errors



Parameterisation – Source Position

★ Basic source parameters

- Position
  - Flux-weighted centroid: $$\boxed{\langle\vec{p}\rangle = \frac{\sum_i \vec{p}_i S(\vec{p}_i)}{\sum_i S(\vec{p}_i)}}$$
  - 3D $\longrightarrow$ $\langle\vec{p}\rangle = \big(\langle x\rangle, \langle y\rangle, \langle z\rangle\big)$
  - Setting $S(\vec{p}_i) = \text{const.}$ will yield geometric centroid
- Important
  - Accurate source mask desirable
  - Negative signals must be excluded
  - Centroid in native pixel coordinates

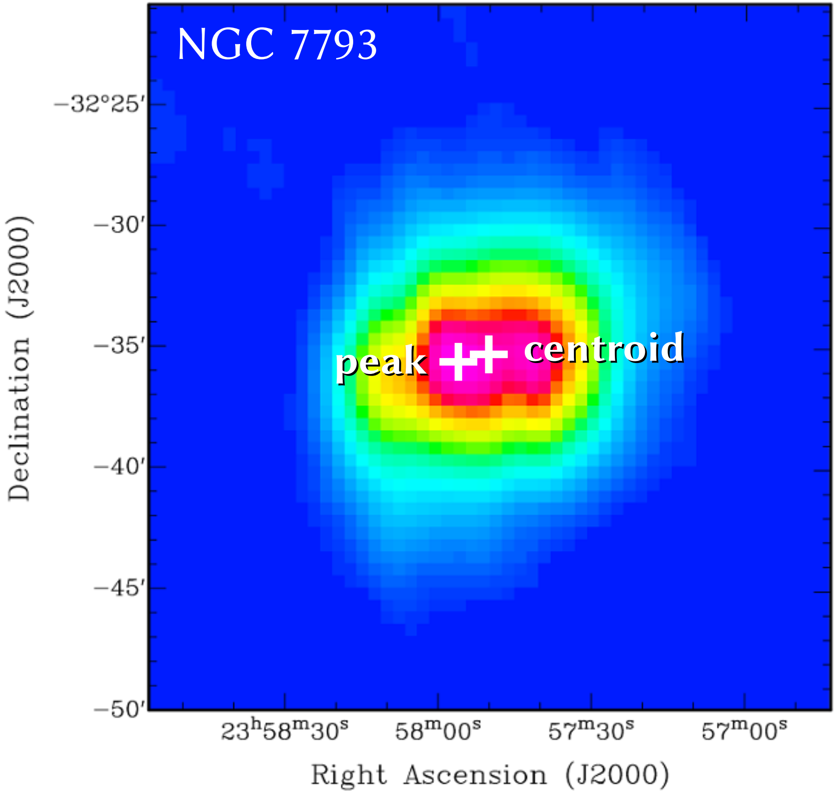
    $\longrightarrow$ Conversion to sky coordinates required

    $\longrightarrow$ World Coordinate System (WCS)

    FITS: https://fits.gsfc.nasa.gov/fits_wcs.html

    wcslib: http://www.atnf.csiro.au/people/mcalabre/WCS/

    Astropy: http://docs.astropy.org/en/stable/wcs/



- Basic source parameters
  - Integrated flux $\boxed{S_{\mathrm{int}} = \dfrac{\Delta z}{\Omega_{\mathrm{PSF}}} \sum_i S(\vec{p}_i)}$
    - Division by beam solid angle required to correct for pixel-to-pixel correlation

$$\Omega_{\mathrm{PSF}} = \frac{\pi\,\theta_a\,\theta_b}{4\ln(2)} \approx 1.133\,\theta_a\,\theta_b$$

for a Gaussian PSF where θ_a, θ_b = FWHM of major, minor axis of beam

    - Units: $\mathrm{Jy\,Hz} = 10^{-26}\ \mathrm{W\,m^{-2}}$ $\longrightarrow$ correct
      $\mathrm{Jy\,km\,s^{-1}}$ $\longrightarrow$ frequently used pseudo-flux unit; better not use
  - H I mass $\boxed{\dfrac{M_{\mathrm{HI}}}{M_{\odot}} = 0.236 \times \dfrac{S_{\mathrm{int}}}{\mathrm{Jy\,km\,s^{-1}}} \times \left(\dfrac{d}{\mathrm{kpc}}\right)^2}$
    - Only valid for optically thin gas at redshift 0

★ Spectral moments

- 0^{th} moment ⟶ Sum of flux densities

$$\boxed{M_0(x,y) = \Delta v \sum_z S(x,y,z)}$$

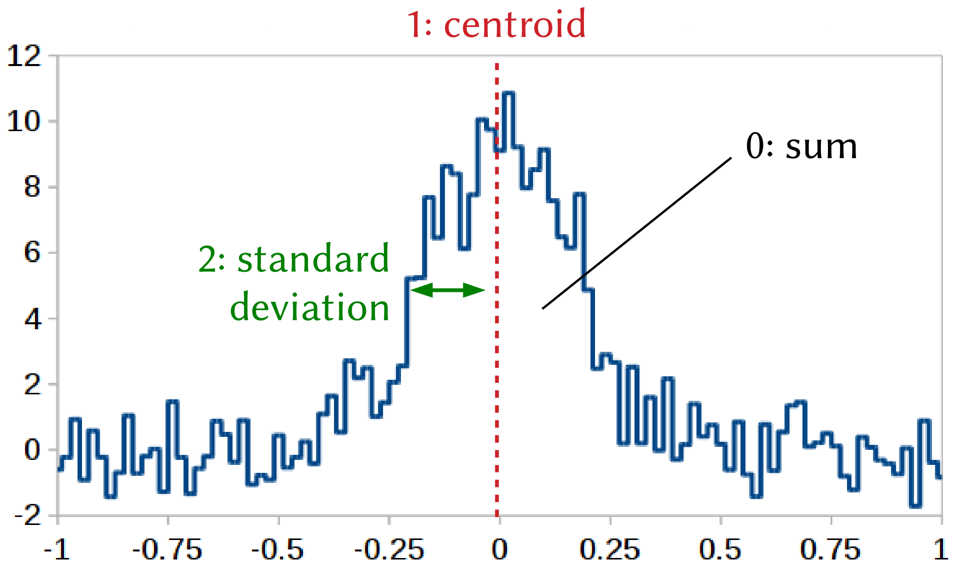
- 1^{st} moment ⟶ Flux-weighted centroid

$$\boxed{M_1(x,y) = \frac{\sum_z v(z)\, S(x,y,z)}{\sum_z S(x,y,z)}}$$

- 2^{nd} moment ⟶ Standard deviation about 1^{st} moment

$$\boxed{M_2(x,y) = \sqrt{\frac{\sum_z [v(z) - M_1(x,y)]^2\, S(x,y,z)}{\sum_z S(x,y,z)}}}$$

- Higher-order moments rarely used
  - 3^{rd} moment (*skewness*), 4^{th} moment (*kurtosis*)

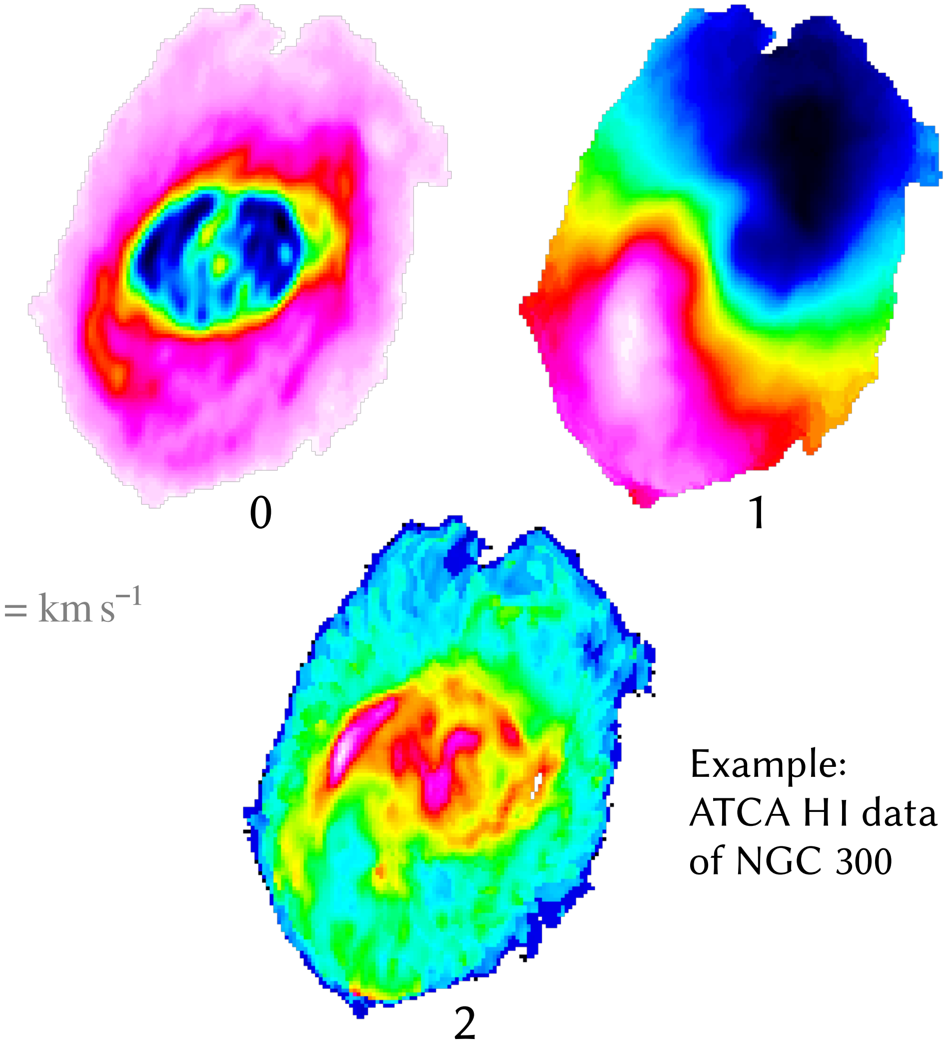


A mask or flux threshold is usually required when calculating moments, as the noise will otherwise dominate the result!

Parameterisation – Spectral Moments

- Spectral moments
  - Units
    - 0^{th} moment
      $\longrightarrow$ Jy Hz or Jy km s^{-1}
      $\longrightarrow$ K Hz or K km s^{-1}
    - 1^{st} and 2^{nd} moments
      $\longrightarrow$ Hz or km s^{-1}
  - 0^{th} moment often converted to H I column density
    - $N_{\text{H I}} = 1.823 \times 10^{18} \int T_{\text{B}} \, \text{d}v$ where $[N_{\text{H I}}] = \text{cm}^{-2}$, $[T_{\text{B}}] = \text{K}$, $[v] = \text{km s}^{-1}$
    - Assumptions
      - Local source at $z = 0$
      - Emission is optically thin
      - Emission is diffuse and fills the telescope beam

Moment analysis is sensitive to noise!



Example:
ATCA H I data
of NGC 300

★ Fitting of spectrum I – *Gaussian Function*

- Useful for fitting and parameterising simple line profiles
- Definition

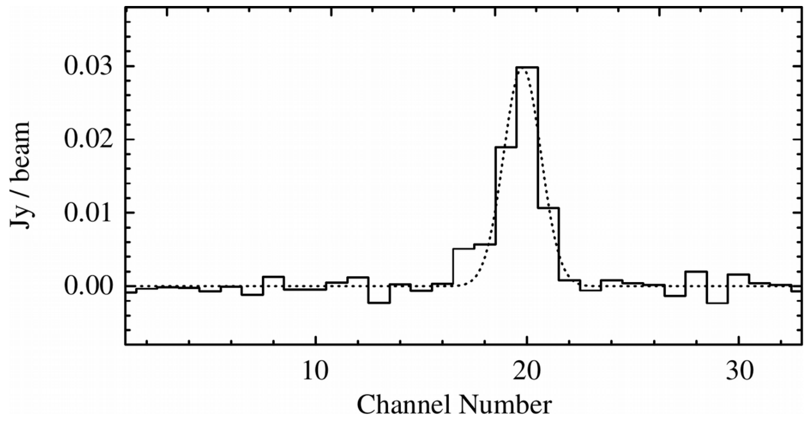
$$\boxed{G(z) = A \exp\left(-\frac{(z - z_0)^2}{2\,\sigma^2}\right)}$$

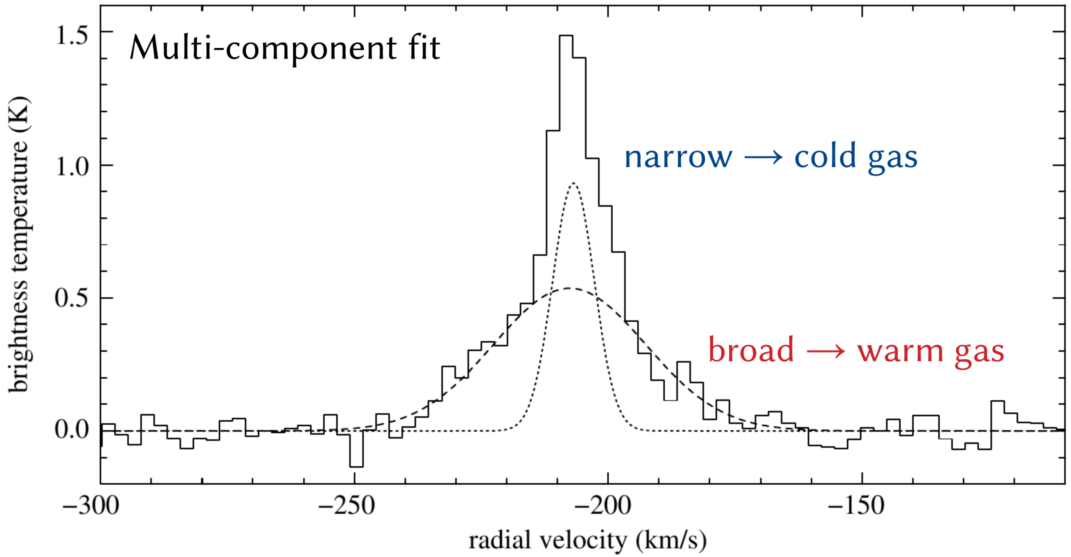
- Relation between w_{50} (FWHM) and σ

$$\boxed{w_{50} = 2\sqrt{2\ln(2)}\,\sigma} \approx 2.3548\,\sigma$$

- Integrated flux

$$\boxed{S_{\rm int} = \int_{-\infty}^{\infty} G(z)\,\mathrm{d}z = \sqrt{2\pi}\,A\,\sigma} \approx 2.5066\,A\,\sigma$$



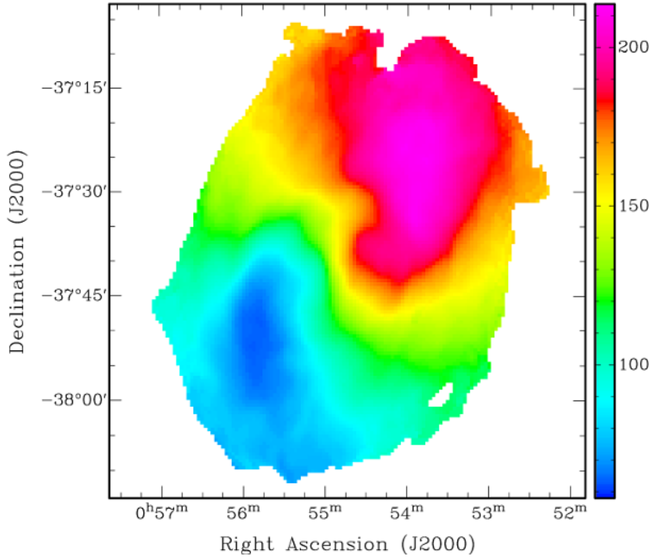


- ★ Fitting of spectrum II – *Gauss–Hermite Polynomial*

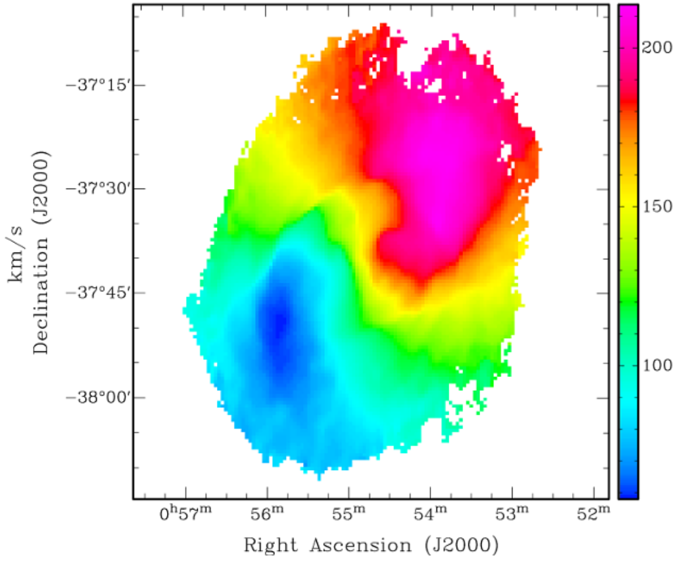
van der Marel & Marijn 1993, ApJ, 407, 525

  - Useful for extracting velocity fields from spatially resolved galaxies for rotation curve analysis
  - Implemented in GIPSY
    - $\phi(x) = a\,e^{-\frac{1}{2}y^2}\left\{1 + \frac{h_3}{\sqrt{6}}(2\sqrt{2}y^3 - 3\sqrt{2}y) + \frac{h_4}{\sqrt{24}}(4y^4 - 12y^2 + 3)\right\} + Z$ where $y \equiv \frac{x-b}{c}$

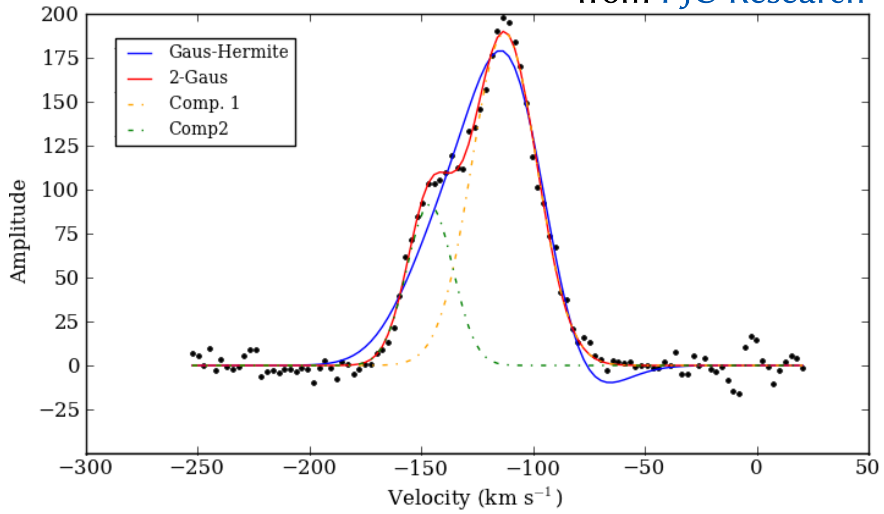
NGC 300 – 1st moment



NGC 300 – Gauss–Hermite



from PJC Research



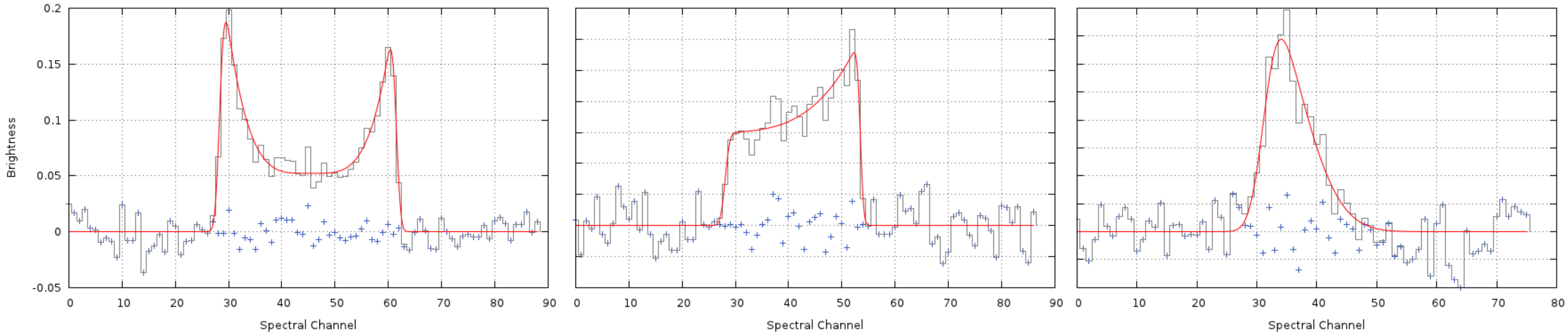
★ Fitting of spectrum III – *Busy Function*

Westmeier et al. 2014, MNRAS, 438, 1176

- Designed to fit double-horn profile of spatially unresolved galaxies
- Product of two error functions and a polynomial

$$B(z) = \frac{a}{4} \times [\mathrm{erf}(b_1(w + z - z_\mathrm{e})) + 1] \times [\mathrm{erf}(b_2(w - z + z_\mathrm{e})) + 1] \times \left[c\left|z - z_\mathrm{p}\right|^n + 1\right]$$

- Software: BusyFit https://github.com/SoFiA-Admin/BusyFit
  BF_dist https://github.com/RussellJurek/busy-function-fitting

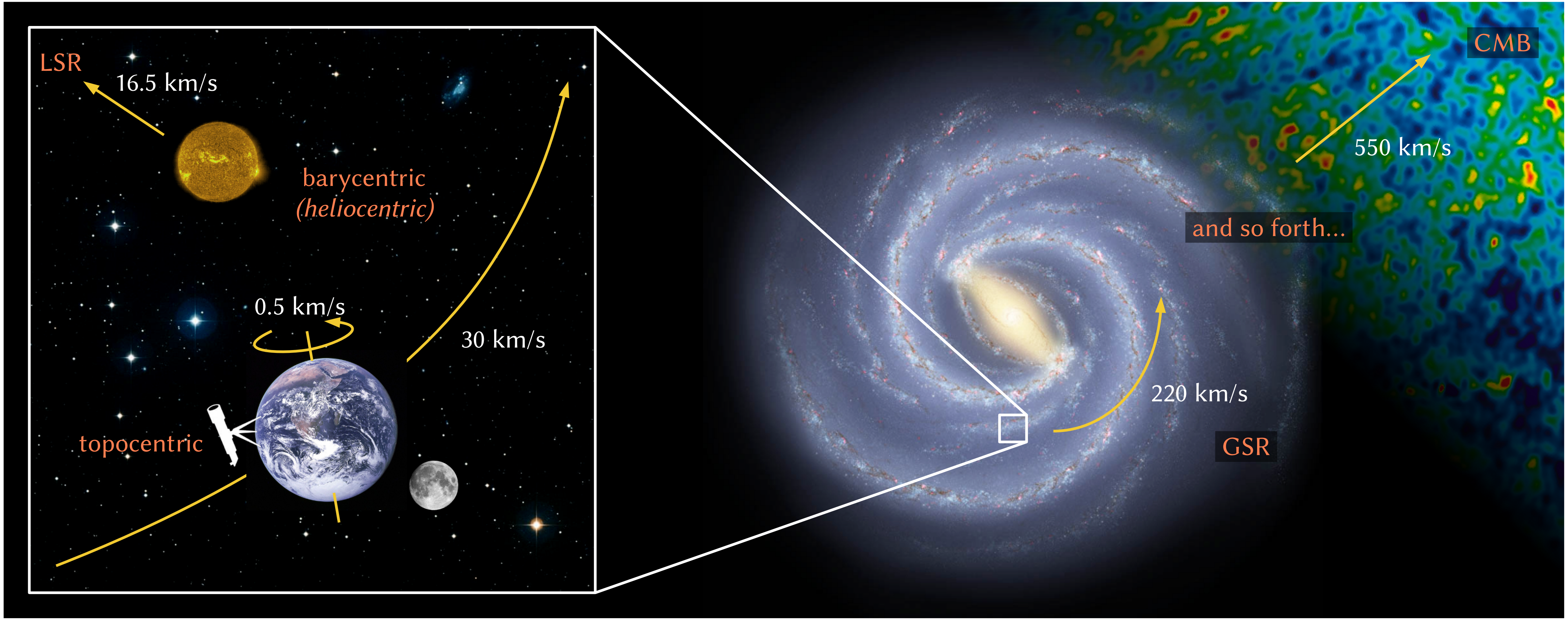


★ Frequency – Redshift – Velocity

- Radio astronomical data cubes usually provided in frequency, f, with constant channel width, Δf
- Relative motion between source and observer
  - Doppler shift between observed frequency, f, and rest frequency, f_0
  - 21-cm H I transition: $f_0 \approx 1.420405751786\ \mathrm{GHz}$
- Reference frames
  - Correction for motion of observer
    - Rotation and orbital motion of the earth
    - Peculiar motion of sun
    - Rotation of Milky Way
    - Motion of Milky Way in Local Group
    - etc.

Parameterisation – Frequency/Velocity/Redshift

★ Velocity rest frames



★ Velocity rest frames

Name	Reference	Description
Topocentric	Observer	Natural rest frame of any observation
Barycentric	Solar System barycentre	Often referred to as "heliocentric"; rest frame most commonly supplied with H I data cubes
Local Standard of Rest (LSR)	Solar neighbourhood	Conversion between barycentric and LSRD: $v_{\mathrm{LSR}} = v_{\mathrm{bar}} + 9\cos(l)\cos(b) + 12\sin(l)\cos(b) + 7\sin(b)$
Galactic Standard of Rest (GSR)	Galactic centre	Conversion between LSRD and GSR: $v_{\mathrm{GSR}} = v_{\mathrm{LSR}} + 220\sin(l)\cos(b)$
LG Standard of Rest (LGSR)	Local Group barycentre	Conversion between GSR and LGSR: $v_{\mathrm{LGSR}} = v_{\mathrm{GSR}} - 88\cos(l)\cos(b) + 64\sin(l)\cos(b) - 43\sin(b)$

These are the rest frames most commonly encountered in radio astronomy.
Anything beyond the barycentric rest frame is inaccurate, in particular the LGSR.

Parameterisation – Frequency/Velocity/Redshift

★ Redshift and velocity

- Definition of redshift: $z \equiv \dfrac{\lambda_{\rm obs} - \lambda_0}{\lambda_0} = \dfrac{f_0 - f_{\rm obs}}{f_{\rm obs}} \quad \Rightarrow \quad \boxed{\dfrac{f_0}{f_{\rm obs}} = 1 + z}$

- Redshift components
  - Cosmological redshift $\longrightarrow$ Hubble expansion of the universe
  - Peculiar redshift $\longrightarrow$ Doppler shift from peculiar velocities
  - Gravitational redshift $\longrightarrow$ GR time dilation in gravitational potential (usually negligible)

- Redshifts are multiplicative
  - $1 + z_{\rm obs} = (1 + z_{\rm cos}) \times (1 + z_{\rm pec}) \times (1 + z_{\rm grav})$

- It is usually not possible to separate redshift components
  - Low redshift $\longrightarrow$ Dominated by peculiar velocities
  - High redshift $\longrightarrow$ Dominated by Hubble expansion

- ★ Peculiar redshift / velocity
  - ▸ Non-relativistic Doppler effect: $\boxed{z_{\mathrm{pec}} = v_{\mathrm{pec}} / \mathrm{c} \equiv \beta}$
    - Valid for small $v_{\mathrm{pec}} \ll \mathrm{c}$
    - Note that generally $\mathrm{c}\, z_{\mathrm{obs}} \neq v$ $\longrightarrow$ "recession velocity" or "optical velocity"
  - ▸ Relativistic Doppler effect: $\boxed{1 + z_{\mathrm{pec}} = \gamma\, [1 + \beta \cos(\vartheta)]}$ where $\gamma \equiv (1 - \beta^2)^{-1/2}$ (Lorentz factor)
    - Depends on transverse velocity!
    - ϑ = angle between direction of motion and line-of-sight from observer to source at time of emission
    - Pure line-of-sight motion: $\boxed{1 + z_{\mathrm{pec}} = \sqrt{\frac{1 + \beta}{1 - \beta}}} \iff \frac{v_{\mathrm{pec}}}{\mathrm{c}} = \frac{f_0^{\,2} - f^{\,2}}{f_0^{\,2} + f^{\,2}}$

Parameterisation – Corrections at Higher Redshift

Don't use "velocity" unless you know what you are doing!

★ Redshift corrections I – Velocity width

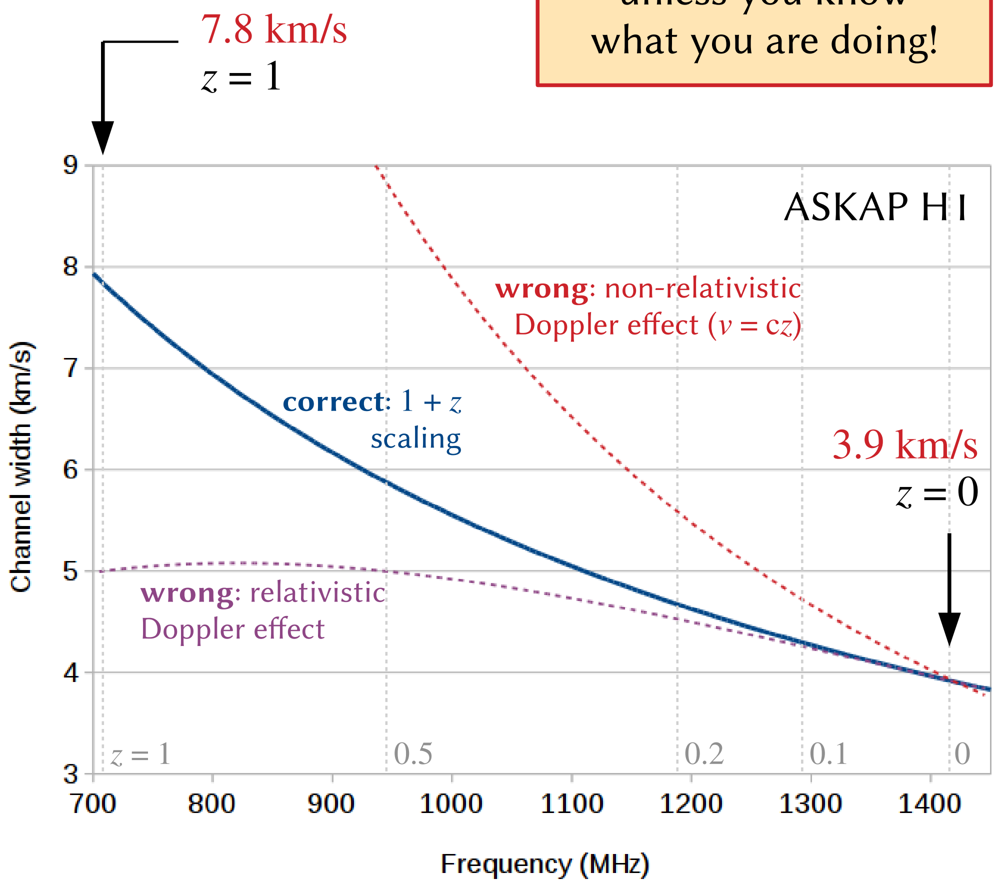
- Assumptions
  - Two objects at same cosmological redshift, z_{cos}
  - Redshift difference, Δz_{obs}, due to velocity difference
- Non-relativistic Doppler effect:

$$z_{\mathrm{pec}} = \beta = \frac{v_{\mathrm{pec}}}{\mathrm{c}}$$

- Peculiar velocity difference along LOS

$$\boxed{\frac{\Delta v_{\mathrm{pec}}}{\mathrm{c}} = \frac{\Delta z_{\mathrm{obs}}}{1 + z_{\mathrm{cos}}}}$$

$$\boxed{\frac{\Delta v_{\mathrm{pec}}}{\mathrm{c}} \approx \frac{1 + z_{\mathrm{cos}}}{f_0} \Delta f_{\mathrm{obs}}}$$



- Redshift corrections II – Flux-related parameters
  - Definition of flux

$$F = \int S \, \mathrm{d}f_{\mathrm{obs}} = \frac{L}{4\pi \, D_{\mathrm{L}}{}^{2}}$$

  - Rayleigh–Jeans law

$$B = \frac{2 \, \mathrm{k_B} \, f^{2} \, T}{\mathrm{c}^{2}} \quad \text{where} \quad I = \frac{S}{\Omega} = \frac{B}{(1+z)^{3}} \quad \text{with telescope beam solid angle } \Omega$$

  - Brightness temperature

Euclidian ($z = 0$): $\boxed{T_{\mathrm{B}} = \frac{\mathrm{c}^{2} \, S}{2 \, \mathrm{k_B} \, f_{0}{}^{2} \, \Omega}}$ Relativistic: $\boxed{T_{\mathrm{B}} = \frac{\mathrm{c}^{2} \, (1+z)^{3} \, S}{2 \, \mathrm{k_B} \, f_{0}{}^{2} \, \Omega}}$

$$\boxed{\frac{T_{\mathrm{B}}}{\mathrm{K}} \approx 6.06 \times 10^{5} \, (1+z)^{3} \, \frac{S}{\mathrm{Jy}} \left(\frac{a \times b}{\mathrm{arcsec}^{2}} \right)^{-1}}$$

- Further information
  - Meyer et al. 2017, PASA, 34, 52

- Redshift corrections II – Flux-related parameters
  - H I column density

$$\boxed{N_{\mathrm{HI}} = \frac{16\pi\,(1+z)^4\,S}{3\,\mathrm{h}\,f_0\,A_{\mathrm{HI}}\,\Omega}}$$

where $A_{\mathrm{HI}} = 2.86888 \times 10^{-15}\ \mathrm{s}^{-1}$ is the spontaneous emission rate of H I

  - Evaluating the constants yields

$$\frac{N_{\mathrm{HI}}}{\mathrm{cm}^{-2}} = 2.64 \times 10^{20}\,(1+z)^4\,\frac{\mathrm{S}}{\mathrm{Jy\,Hz}}\left(\frac{\Omega}{\mathrm{arcsec}^2}\right)^{-1}$$

$$= 2.33 \times 10^{20}\,(1+z)^4\,\frac{\mathrm{S}}{\mathrm{Jy\,Hz}}\left(\frac{a \times b}{\mathrm{arcsec}^2}\right)^{-1}$$

for a Gaussian beam

- Further information
  - Meyer et al. 2017, PASA, 34, 52

- Redshift corrections II – Flux-related parameters
  - H I mass

$$\boxed{M_{\mathrm{HI}} = \frac{16\pi\, m_{\mathrm{H}}\, D_{\mathrm{L}}{}^{2}\, S}{3\,\mathrm{h}\, f_{0}\, A_{\mathrm{HI}}}}$$

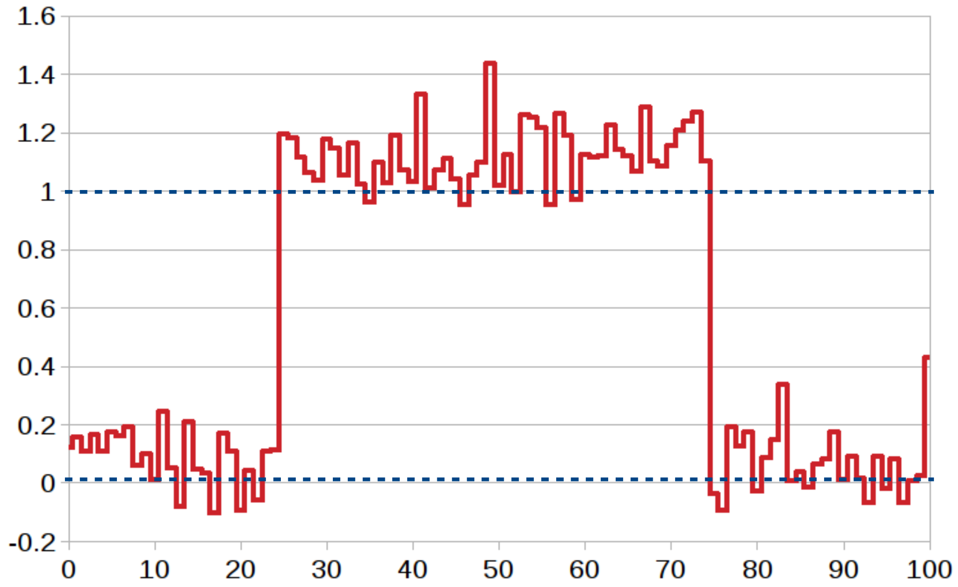
  - where
    - $A_{\mathrm{HI}} = 2.86888 \times 10^{-15}\ \mathrm{s}^{-1}$ is the spontaneous emission rate of H I
    - $m_{\mathrm{H}} = 1.673533 \times 10^{-27}$ kg is the mass of a hydrogen atom
    - $D_{\mathrm{L}}(z)$ is the redshift- and cosmology-dependent luminosity distance
  - H I mass depends on assumptions about cosmology

- Further information
  - Meyer et al. 2017, PASA, 34, 52

- Uncertainties
  - Measurement errors usually dominated by systematic errors
    - flux calibration
    - continuum subtraction
    - spectral bandpass calibration
    - image deconvolution
    - radio frequency interference
    - missing diffuse flux (due to lack of short spacings)
    - parameterisation errors due to insufficient source mask
    - source confusion (multiple sources perceived as one)
    - systematic errors in source distance measurements
    - ...

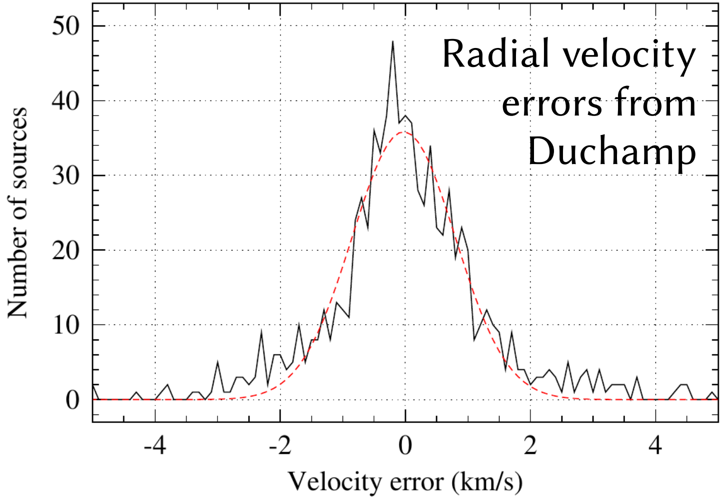
Parameterisation – Uncertainties

- ★ Example
  - Source with $S = 1$ Jy over $N = 50$ channels
  - Noise level of $\sigma = 0.1$ Jy
  - Flux calibration error of 5%
  - Bandpass error of 0.1 Jy
- ★ True flux and statistical uncertainty
  - $F_{\text{true}} = 50\ \text{Jy}, \ \sigma_{\text{stat}} = \sigma \times \sqrt{N} \approx 0.7\ \text{Jy}$
- ★ Measured flux
  - $F_{\text{meas}} = 57.5 \pm 0.7\ \text{Jy}$ (15.5% too high)
- ★ Discrepancy
  - $(F_{\text{meas}} - F_{\text{true}}) / \sigma_{\text{stat}} \approx 10.6$

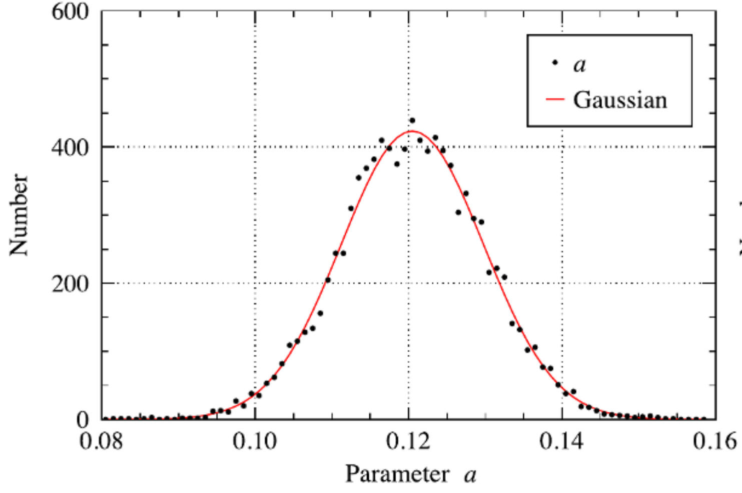


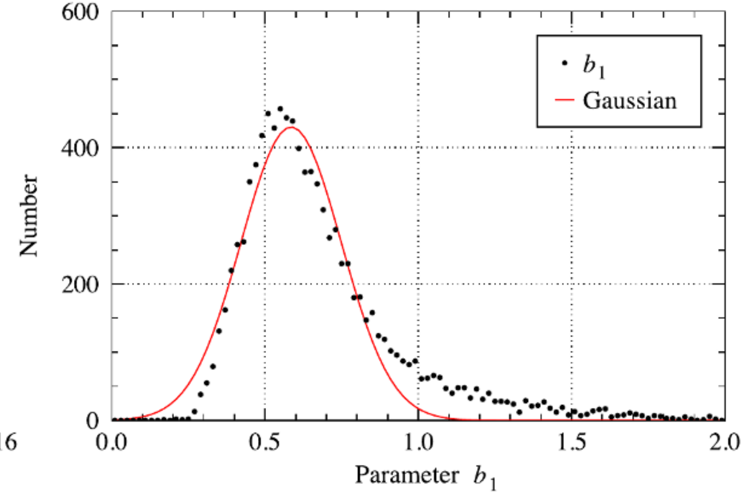
Parameterisation – Uncertainties

- How to get realistic error estimates?
  - Numerical methods
  - Common techniques
    - Injection of artificial sources into data
    - Shifting of source mask to "empty" regions of data cube



- Additional problem
  - Errors may not be Gaussian
  - Mean & standard deviation ($\mu \pm \sigma$) no longer meaningful
    - Numerical error analysis required
  - Example
    - Busy Function
    - a is Gaussian, but not b_1





Summary

★ Key points to take away

- Source finding is non-trivial and needs fine-tuning
- Optimal convolution filters required to detect sources
- Compromise between high completeness and high reliability
  - Reliability calculation can help, but clean data required
- Accurate source masks required for parameterisation
  - Beware of biases
- Difference between observed frequency/redshift and source-frame velocity
- Velocity resolution changes with redshift
  - Corrections required beyond redshift 0
- Distance-dependent parameters (e.g. H I mass) are cosmology-dependent
- Parameterisation errors usually dominated by systematic errors
  - Numerical error analysis required