Source extraction and characterisation I – continuum

Dr. Paul Hancock

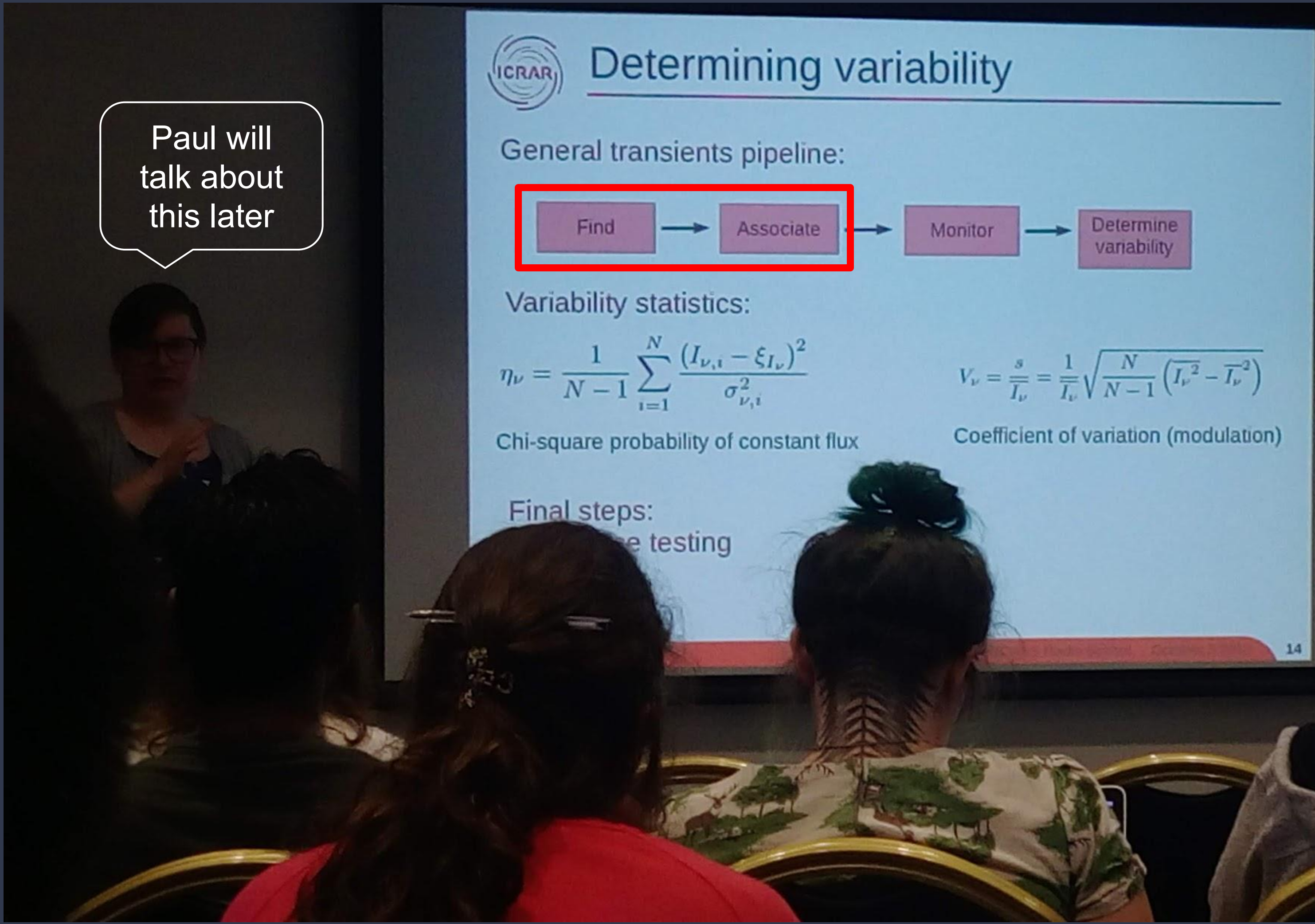
paul.hancock@curtin.edu.au
@drpaulhancock

Curtin University

MWA
MURCHISON
WIDEFIELD
ARRAY

ICRAR
International Centre for Radio Astronomy Research

Intro



Paul will talk about this later
ICRAR
Determining variability
General transients pipeline:
Find
Associate
Monitor
Determine variability
Variability statistics:
$\eta_\nu = \frac{1}{N-1}\sum_{i=1}^{N}\frac{(I_{\nu,i}-\xi_{I_\nu})^2}{\sigma_{\nu,i}^2}$
$V_\nu = \frac{s}{\overline{I_\nu}} = \frac{1}{\overline{I_\nu}}\sqrt{\frac{N}{N-1}\left(\overline{I_\nu^2}-\overline{I_\nu}^2\right)}$
Chi-square probability of constant flux
Coefficient of variation (modulation)
Final steps:
14

Everything you do wrong looks like variability

If you care about variability, then you care about all the ways that things can go wrong.

Eg the presence or changes in:

- Observing conditions
- RFI
- Calibration
- Imaging
- Detection of features
- Characterisation of features
- Analysis and methodology
- Work-flows

Variability can be:

1. Astrophysical
   a. Intrinsic (SNe)
   b. Extrinsic (scintillation)
2. Environmental (RFI, the ionosphere)
3. Instrumental (gain, bandpass stability)
4. Methodological (dodgy math!)

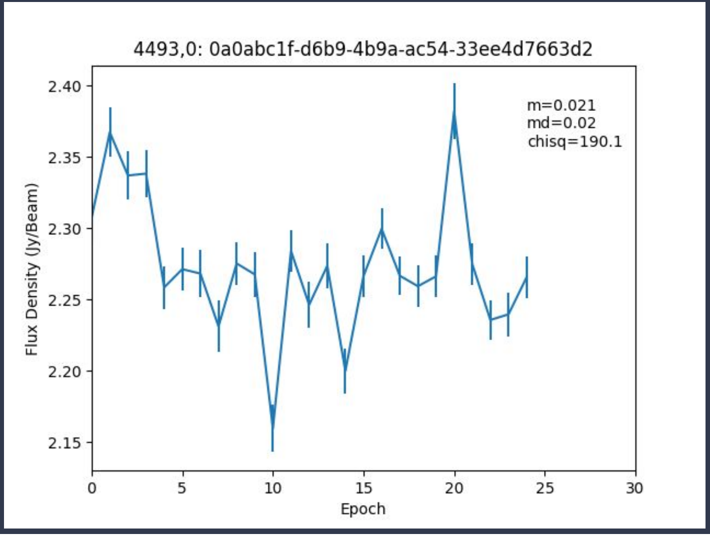
Measuring Variability

Problems:

- Masked/missing data points
- Upper/lower limits
- Non-uniform uncertainties
- Inaccurate uncertainties
- Separating significance and degree

Solutions:

- Better source characterisation
- Better statistical models

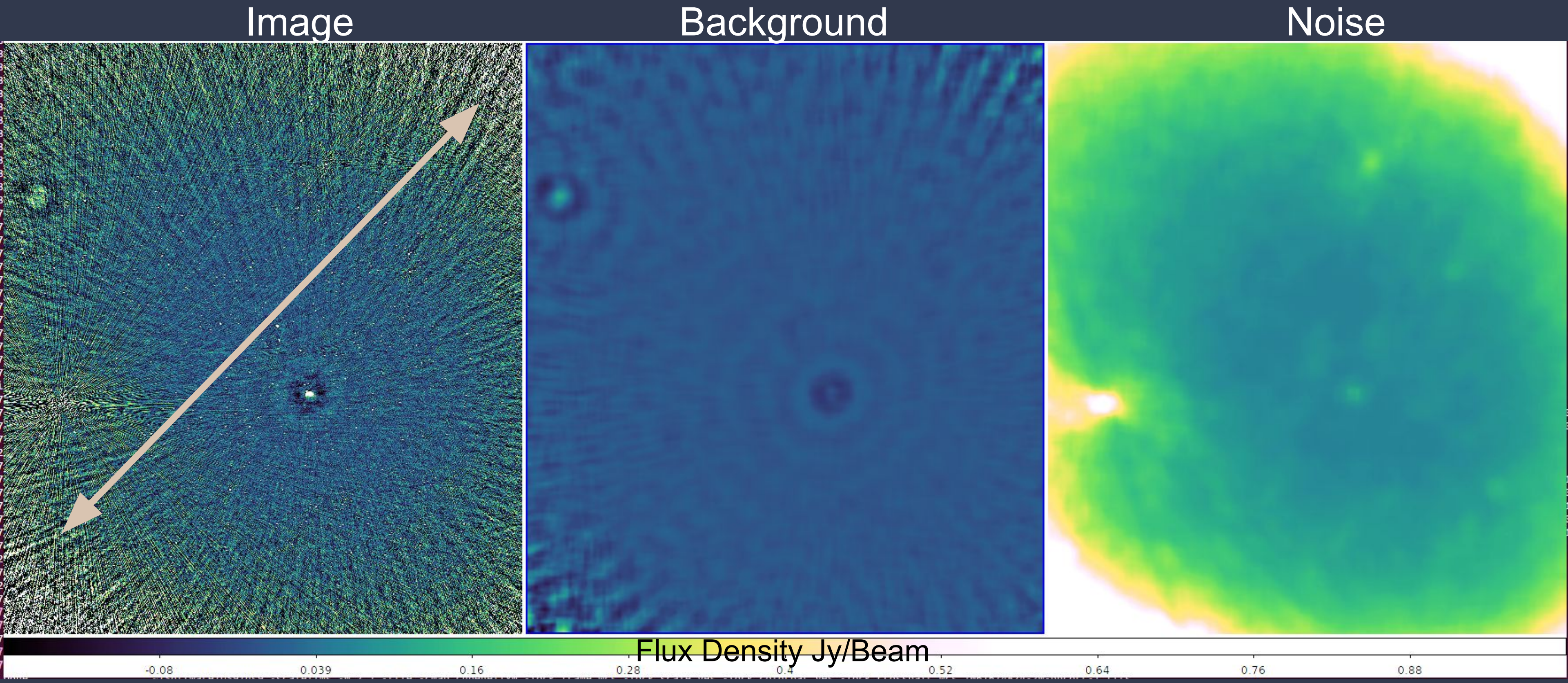


Source Finding Done Right

Assumptions:

- Compact sources
- Continuum images

Snapshot Image: Data model



Zoom

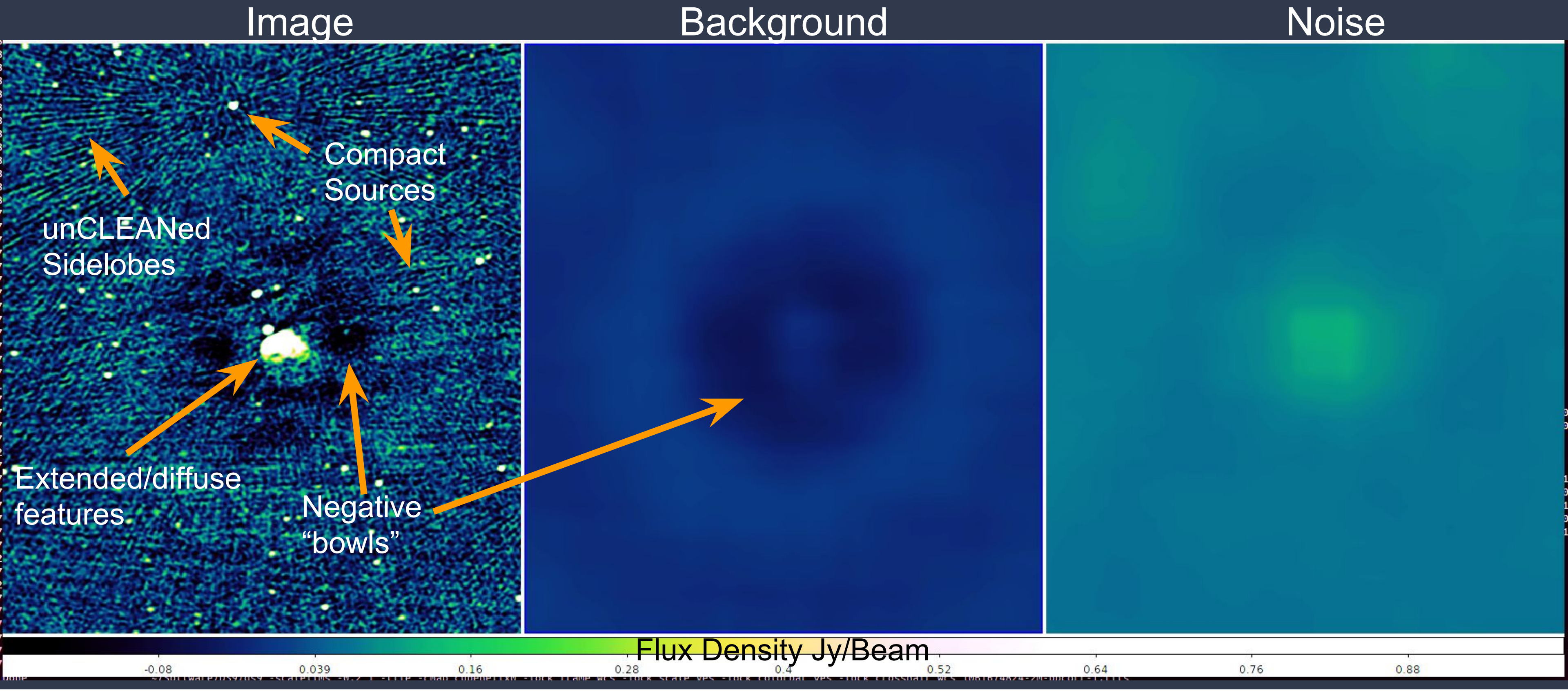
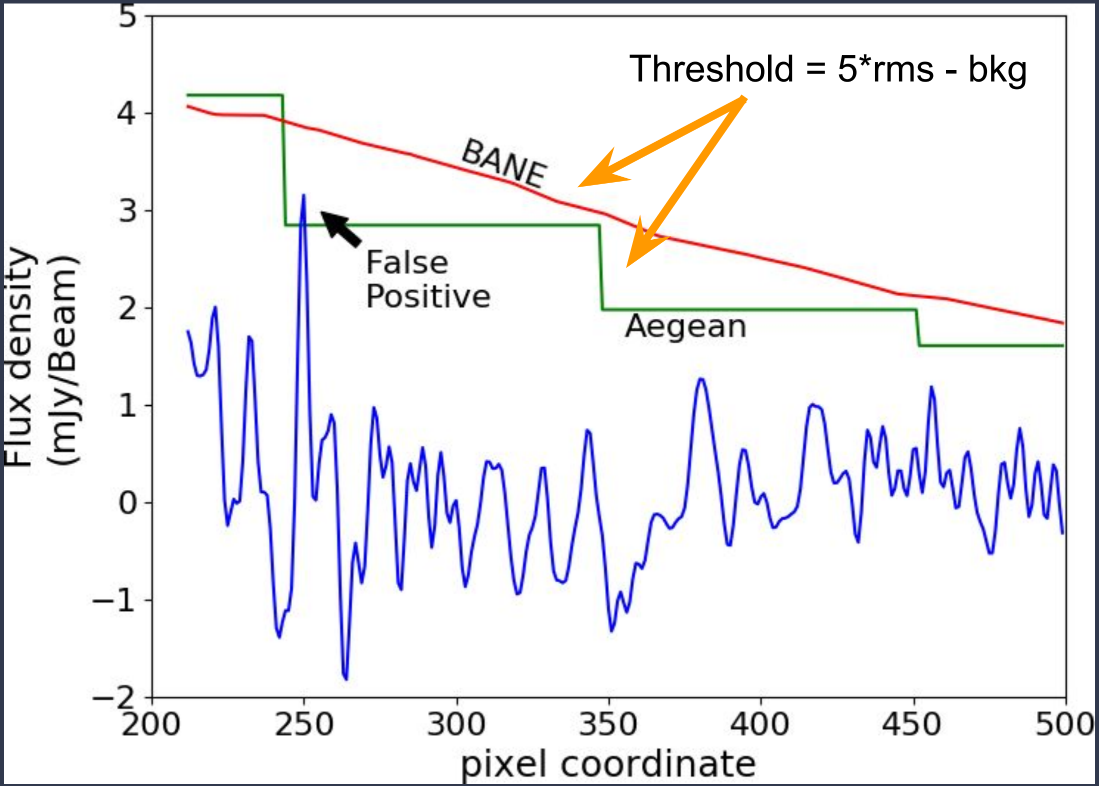
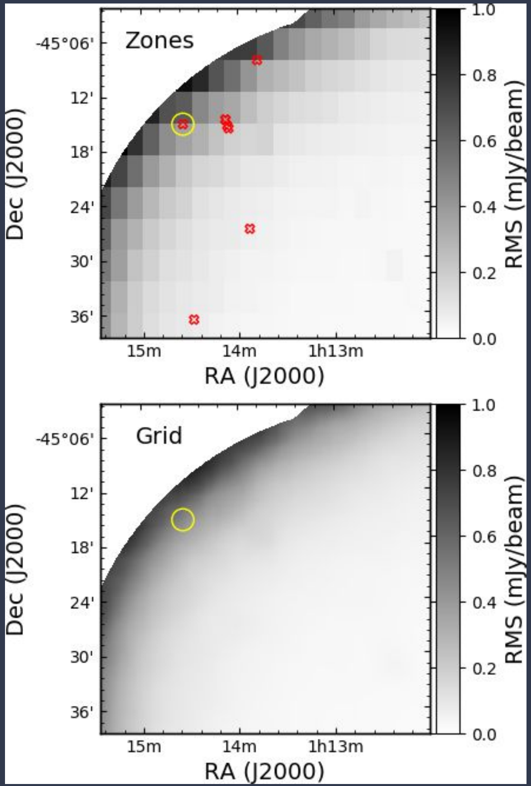


Image
Background
Noise
Compact
Sources
unCLEANed
Sidelobes
Extended/diffuse
features
Negative
“bowls”
Flux Density Jy/Beam
-0.08
0.039
0.16
0.28
0.4
0.52
0.64
0.76
0.88

Finding sources – thresholding

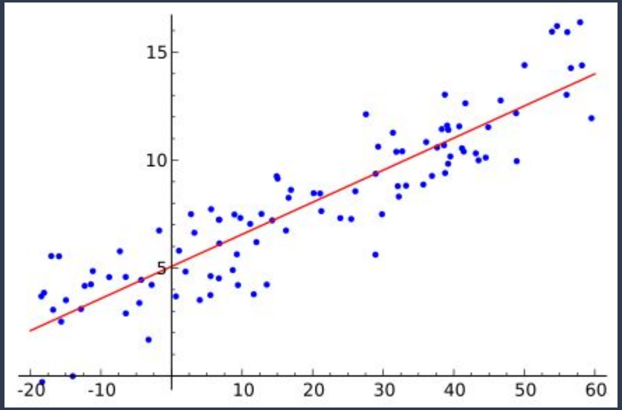




(linear) Least squares fitting

Given:

- x - data
- f(θ;x) - model data with parameters θ



commons.wikimedia.org

Goal:

- Minimise the sum of the square of the residuals

$$\arg \mathrm{Min} \sum \left(f(\theta;x) - x\right)^2$$

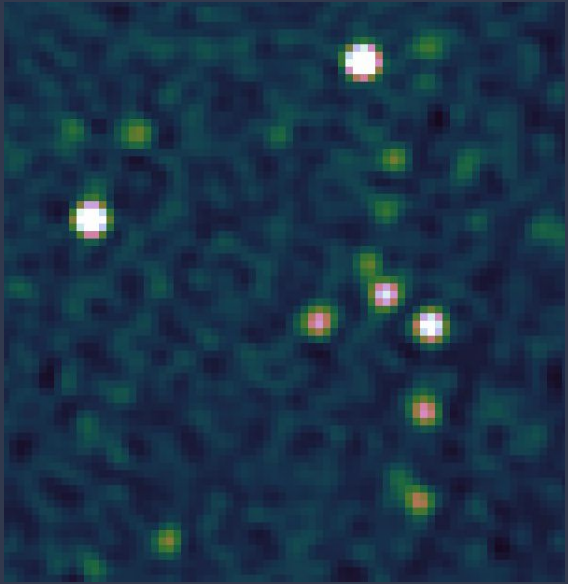
- a.k.a χ^2 minimisation

For linear models and data that is independent and identically distributed, least squares minimisation is unbiased, and has minimum variance.

Radio Images

We fit with a source model that is Gaussian

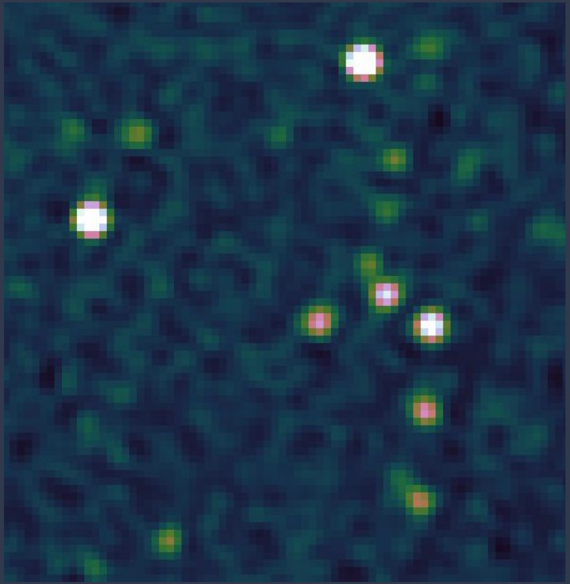
$$f(x, y) = Ae^{-\left(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right)}$$



Radio Images

We fit with a source model that is Gaussian

$$f(x,y) = A\boxed{e}^{-\left(\frac{(x-x_0)^{\boxed{2}}}{2\sigma_x^2}+\frac{(y-y_0)^{\boxed{2}}}{2\sigma_y^2}\right)}$$

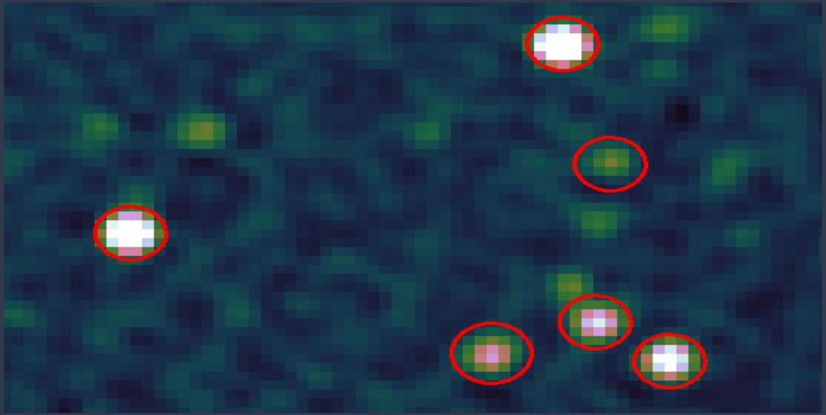


Not linear, not even close

(non linear) Least squares fitting

Given:

- x - data
- f(θ;x) - model data with parameters θ



Goal:

- Minimise the sum of the square of the residuals

$$\arg\,\mathrm{Min} \sum \left(f(\theta; x) - x\right)^2$$

- a.k.a χ^2 minimisation

For non linear models least squares minimisation gives a biased result.

All parameters are biased, even the ‘linear ones’ like amplitude

Quantifying Bias

Refreiger & Brown 1998 (arXiv:9803279) describe the expected bias as:

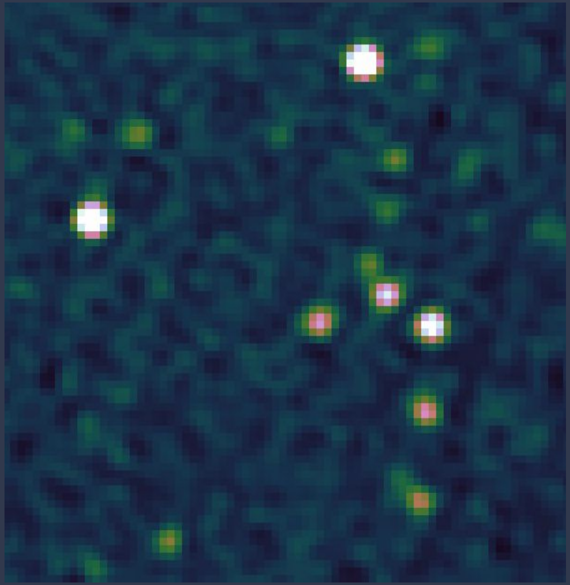
$$\langle a_i \rangle = \hat{a}_i \boxed{- \frac{1}{2}\sigma_N^2 B_{lkj} D_{li} D_{kj}} + O(\mathrm{SNR}_s^{-3})$$

Where

$$\begin{aligned} D_{ij} &= (H^{-1})_{ij}, \\ H_{ij} &= \sum_p \frac{\partial F}{\partial a_i}(\mathbf{x}^p; \hat{\mathbf{a}}) \frac{\partial F}{\partial a_j}(\mathbf{x}^p; \hat{\mathbf{a}}), \\ B_{ijk} &= \sum_p \frac{\partial F}{\partial a_i}(\mathbf{x}^p; \hat{\mathbf{a}}) \frac{\partial^2 F}{\partial a_j \partial a_k}(\mathbf{x}^p; \hat{\mathbf{a}}), \end{aligned}$$

* Math is for demonstration purposes only - Do not try this at home

Radio Images Again



Data are correlated:

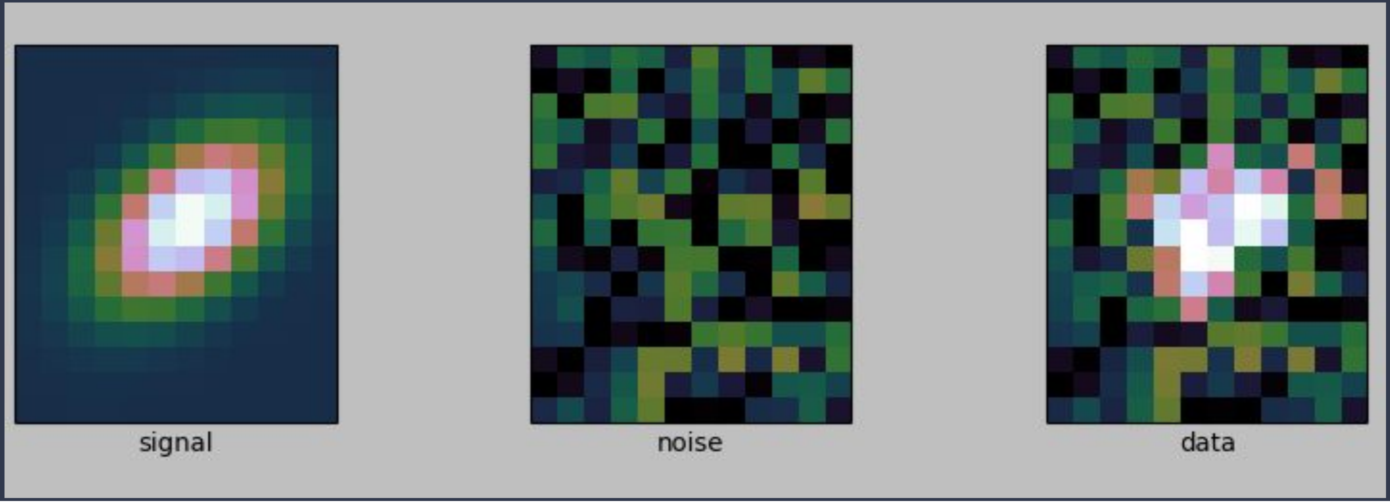
corr(x,y) = Dirty Beam / Point Spread Function

Even worse:

CLEAN-ing modifies the correlation function

Our data

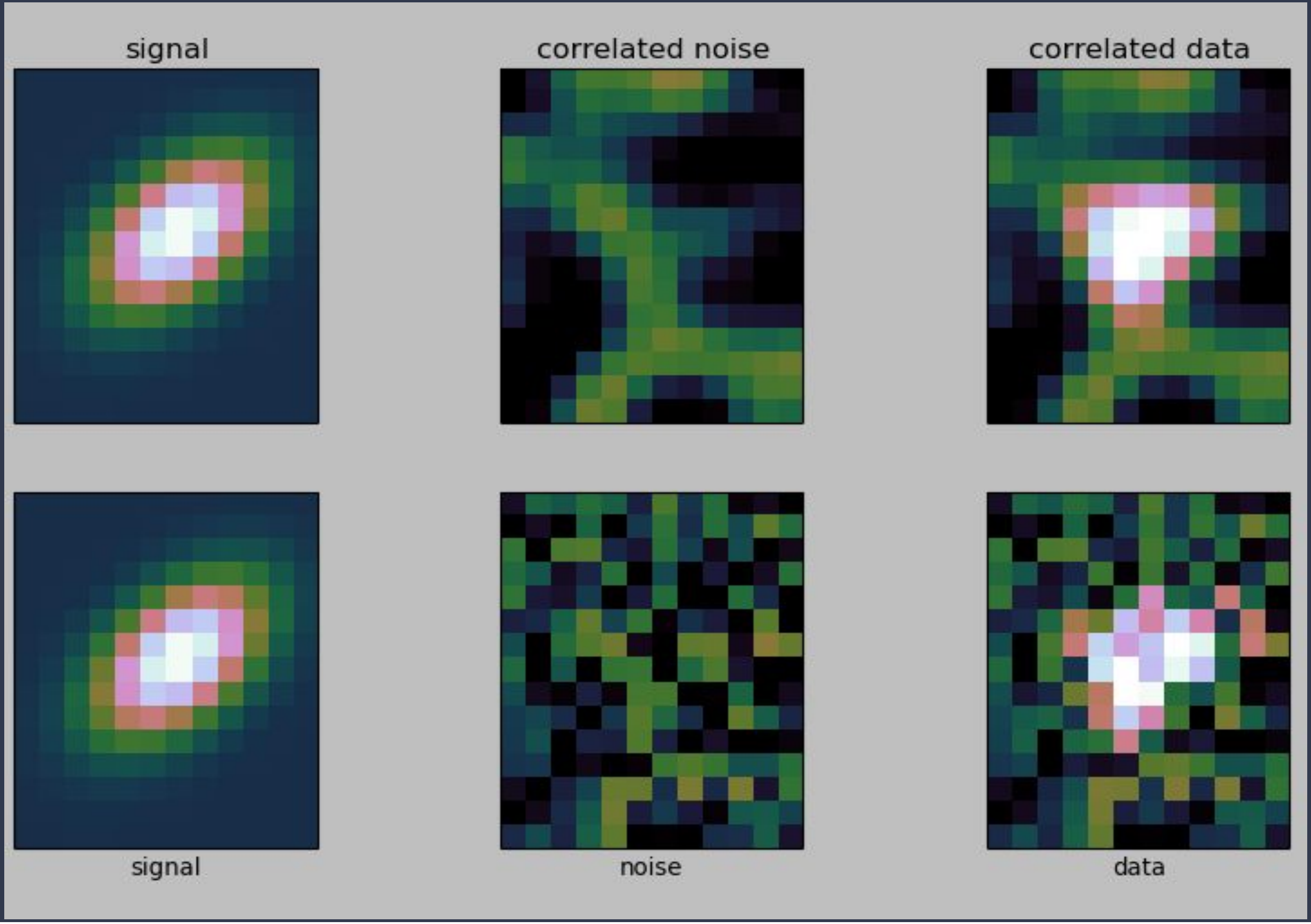
What our fitting algorithms assume we have



Our data

What we **actually** have

What our fitting algorithms **assume** we have



Correlated Data

Increases bias in all parameters

Additional bias towards local noise peaks at low SNR

Nearby sources now have correlated parameters

Approaches

- Ignore correlations completely
- Fit as usual and then fiddle errors to account for correlations
  - How many DoF do we “really” have?
- Fit as usual and then replace errors with empirically derived quantities
  - Condon 1997
- Fit as usual and report errors based on analytical solution
  - Refreiger & Brown 1998 (arXiv)
- Fit with a cost function that incorporates a correction for correlated data
  - Aegean 2.0, Hancock et al. 2018

How do we do better?

Given:

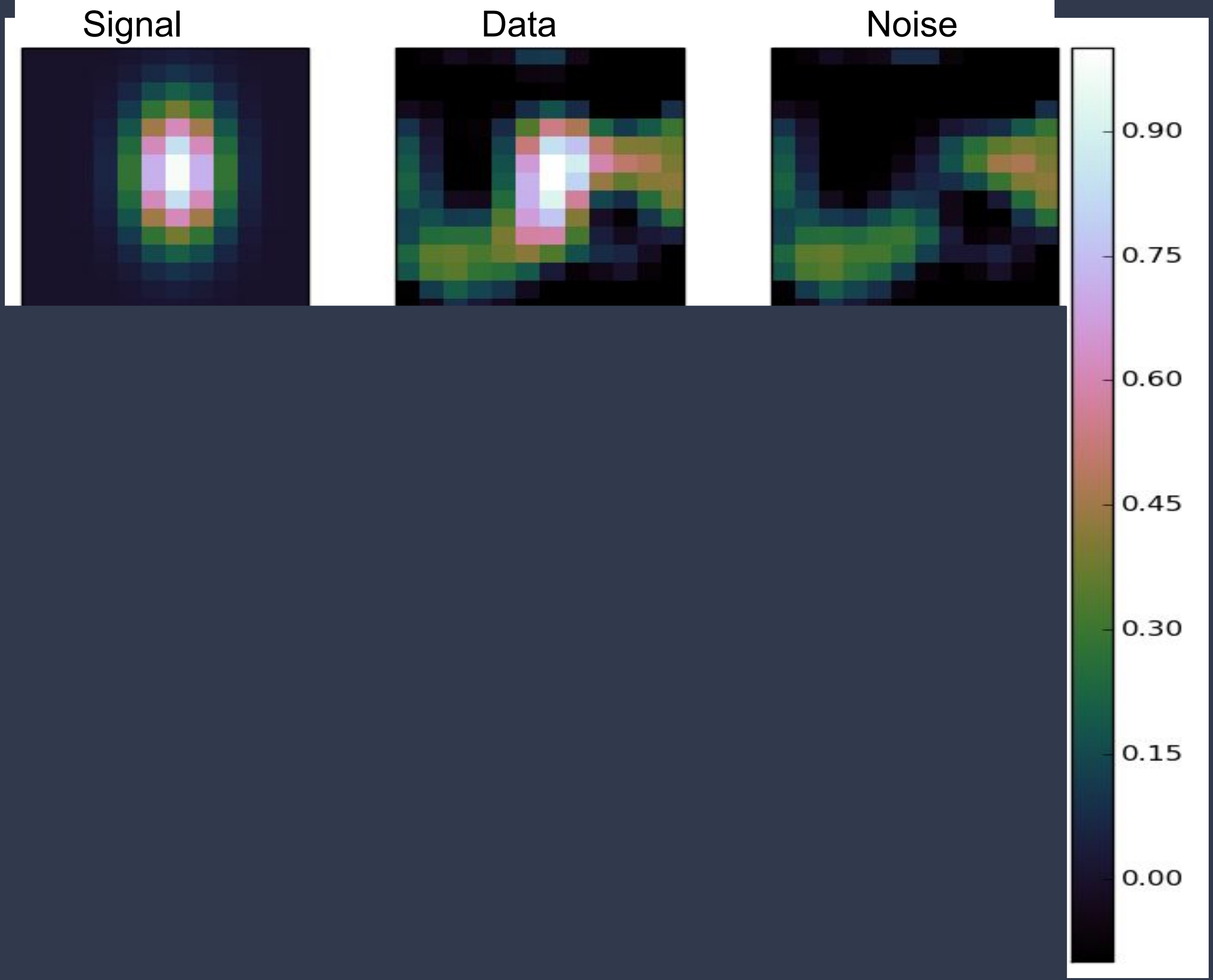
- x - data
- $f(\theta;x)$ - model data with parameters θ
- Covariance matrix C

Goal:

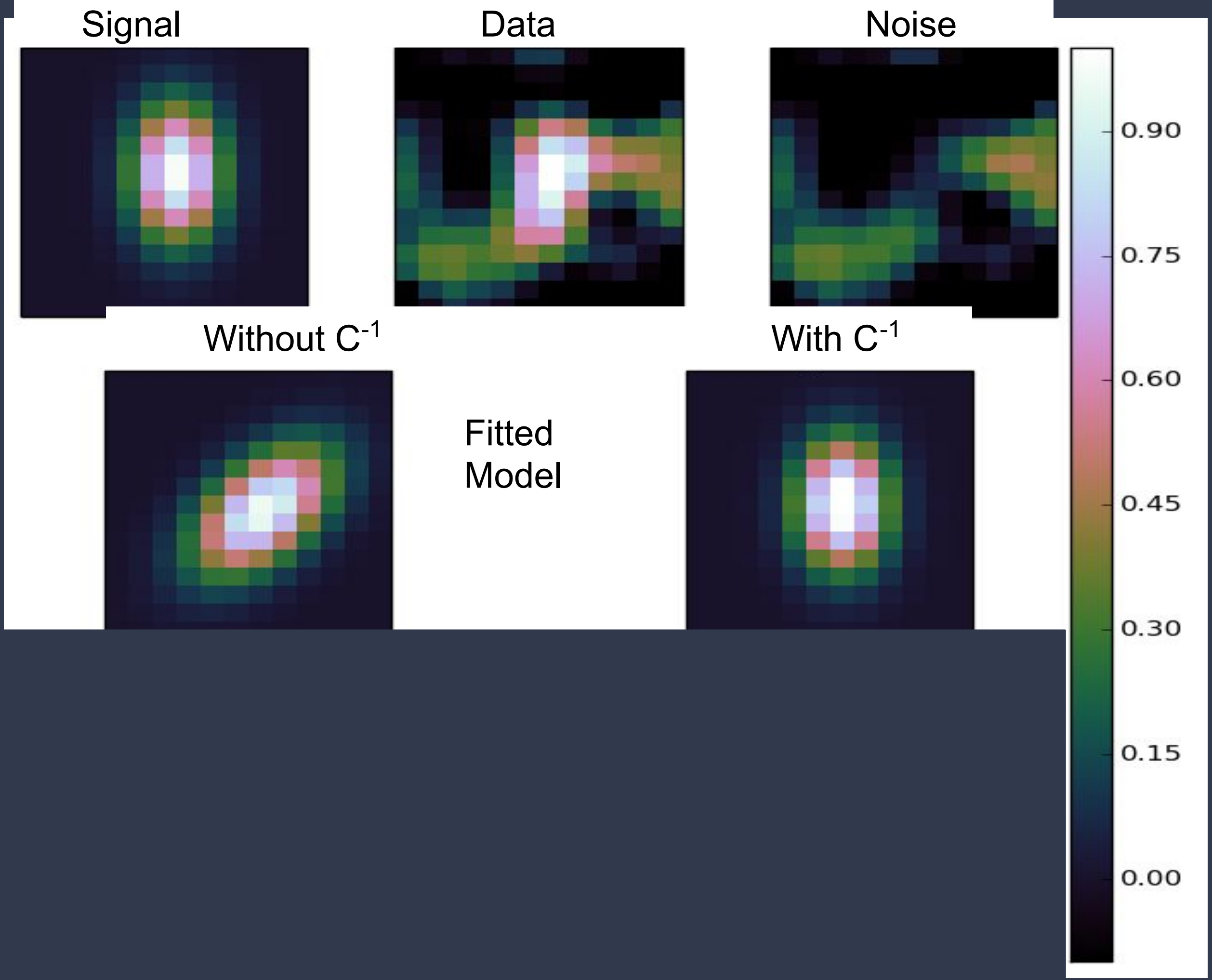
- Minimise the sum of the square of the residuals modified by the inverse covariance matrix

$$\text{Min} \{ (f(\theta;x) - x)^T C^{-1} (f(\theta;x) - x) \}$$

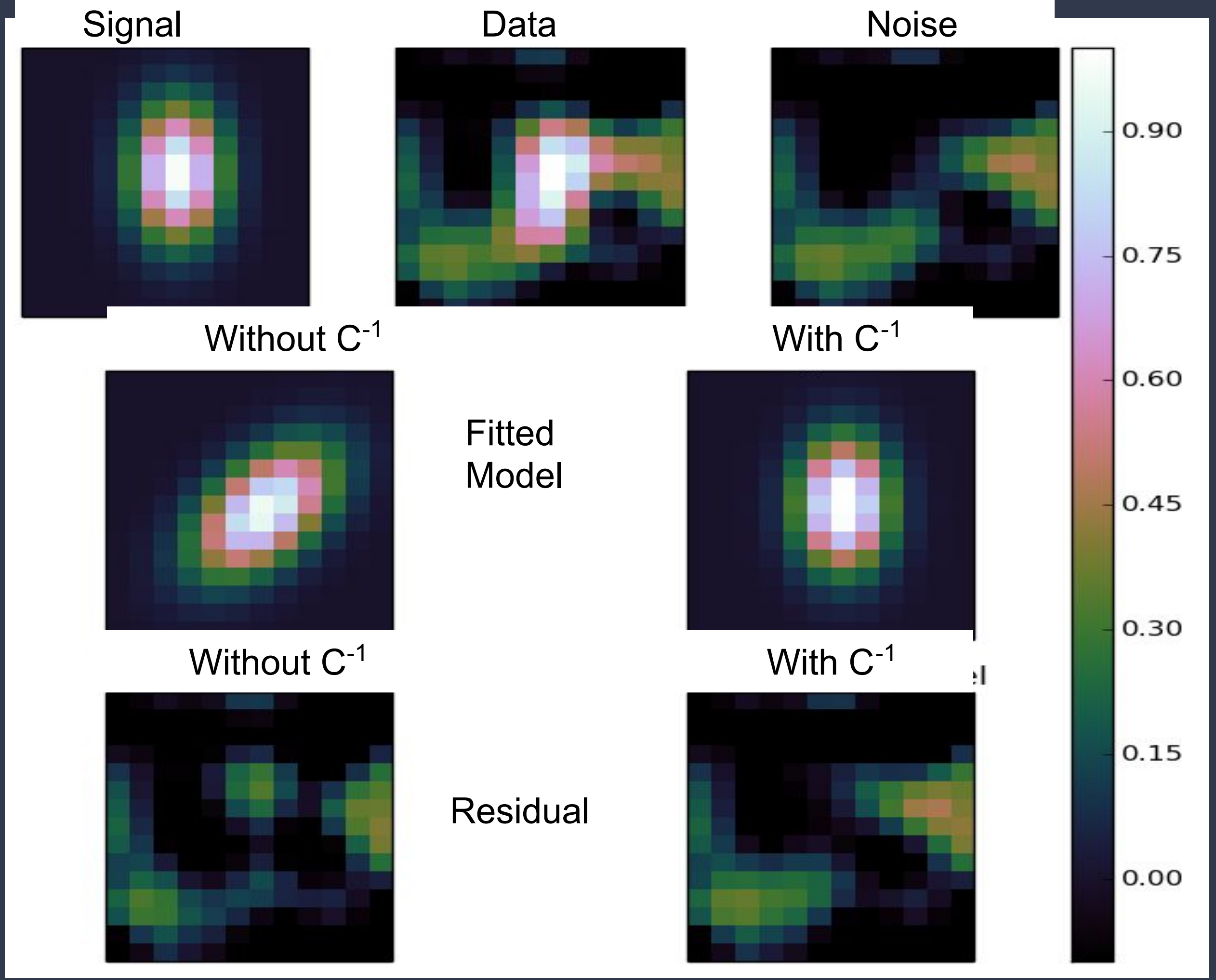
Fitting with C^{-1}



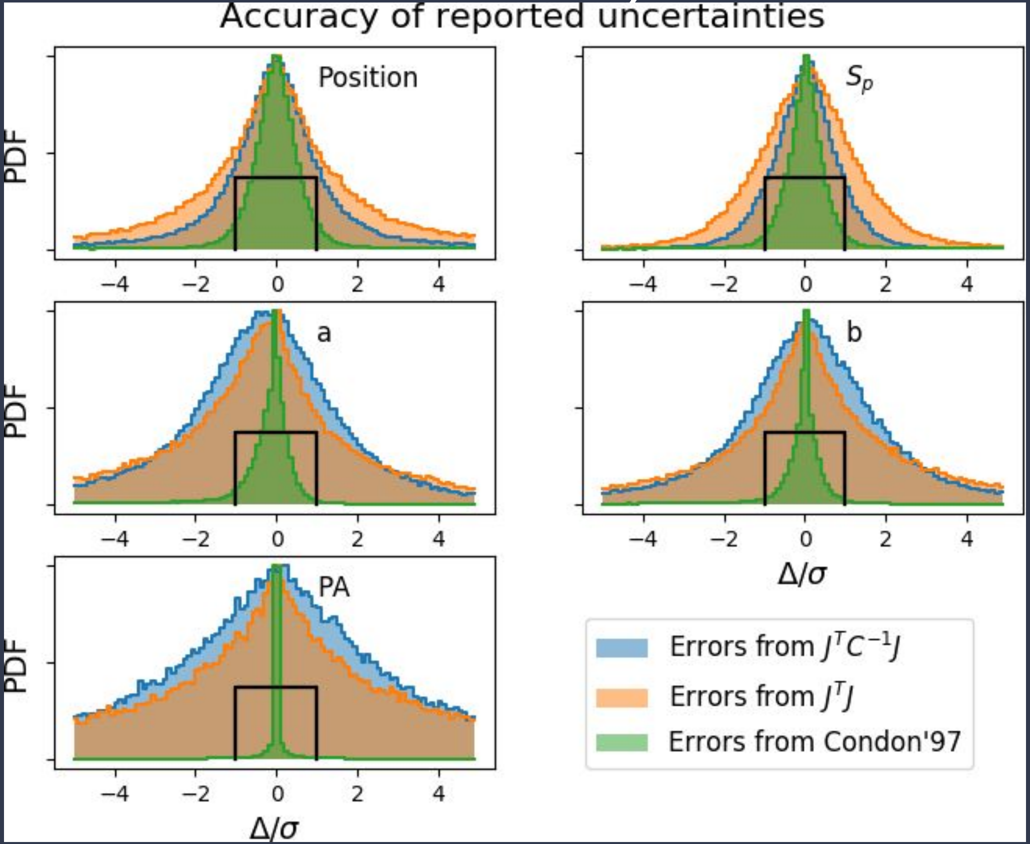
Fitting with C^{-1}

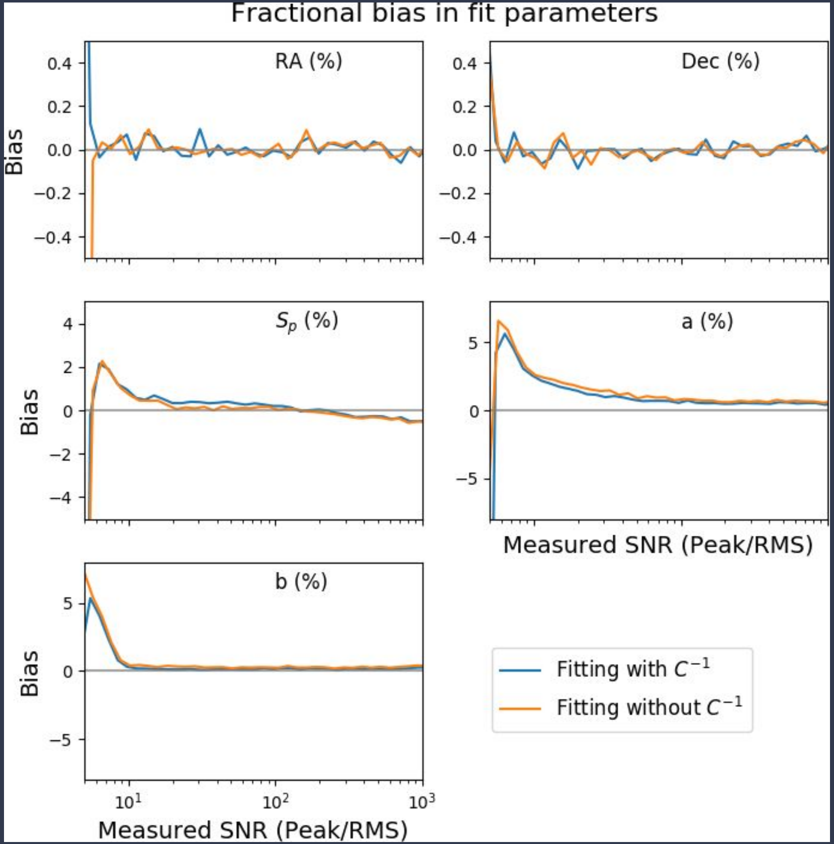


Fitting with C^{-1}



Accurate errors, biases

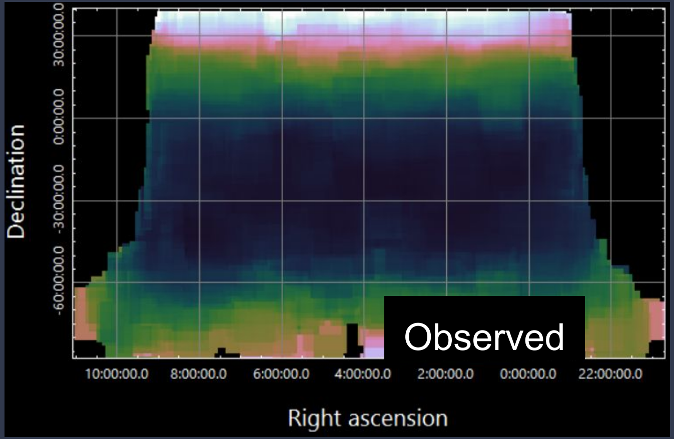


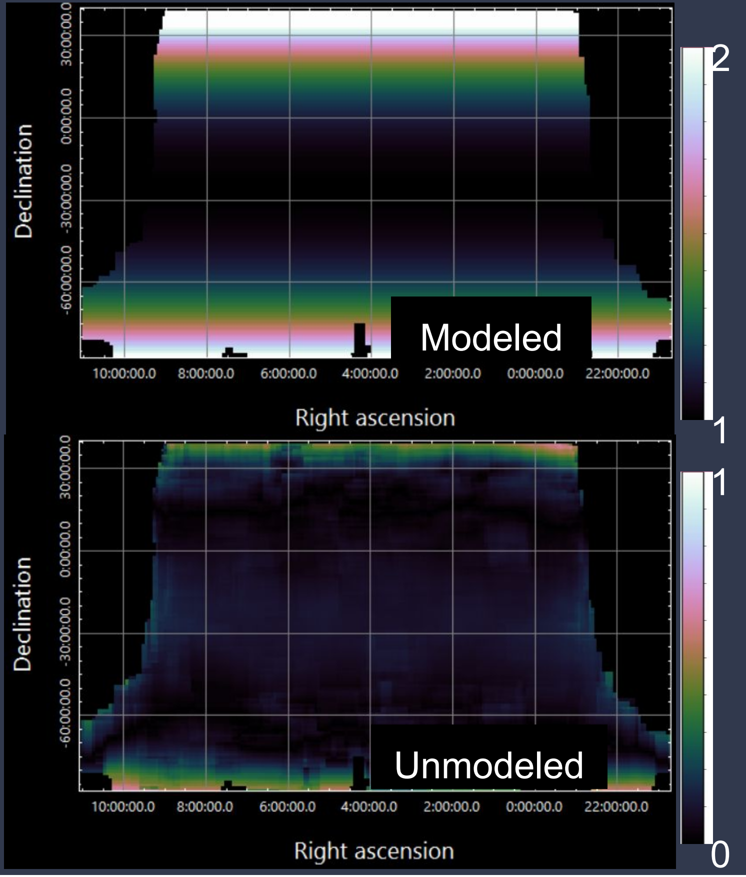


Making

- Catalogues
- Light-curves
- SEDs

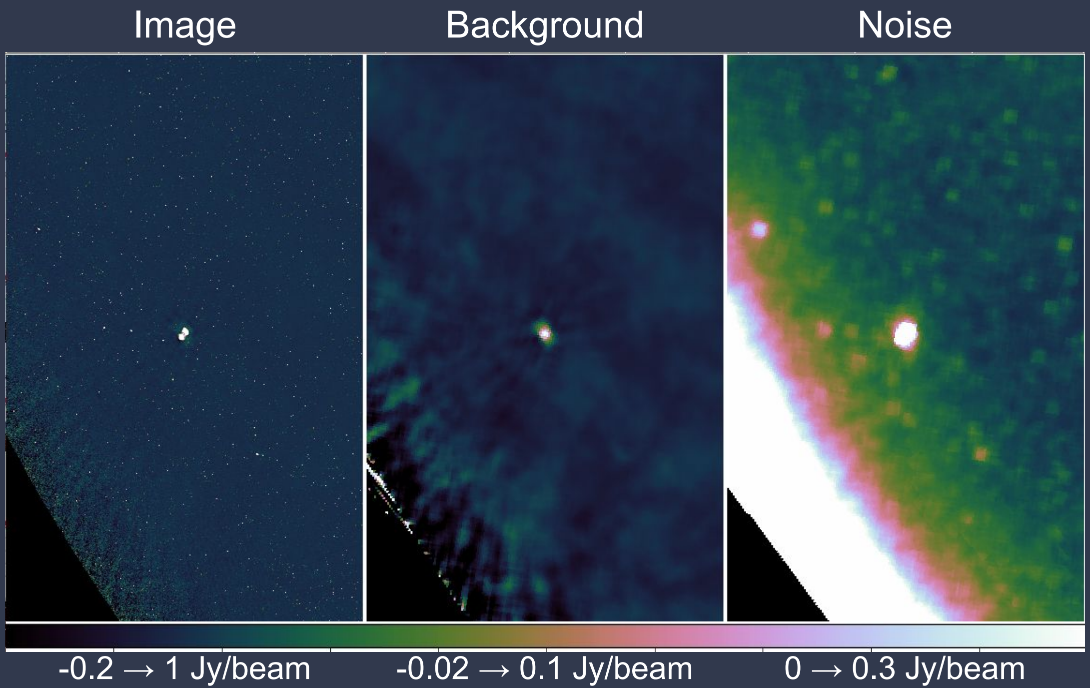
Direction Dependant Synthesized Beam

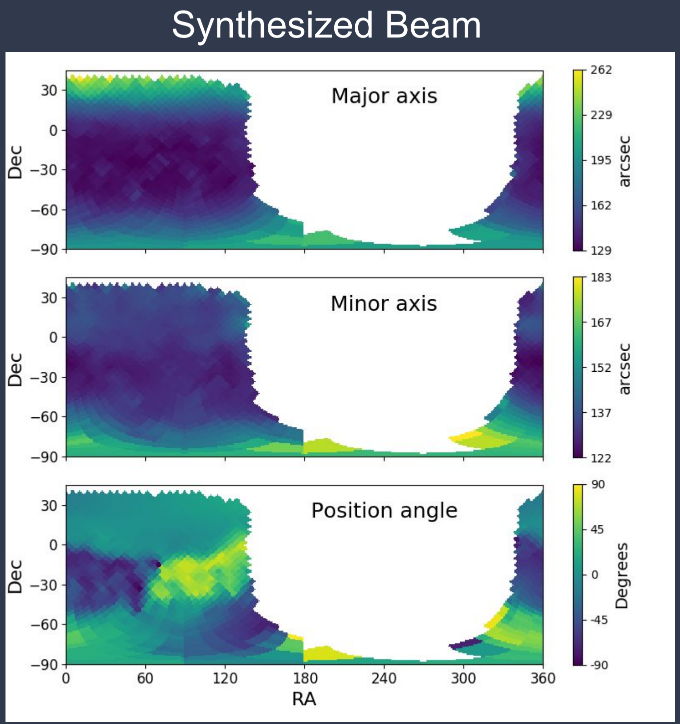




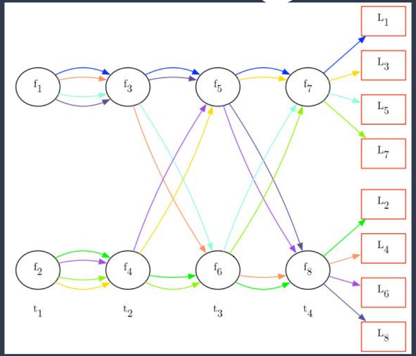
Major Axis
Size Relative
to Zenith

Catalogues at large FoV



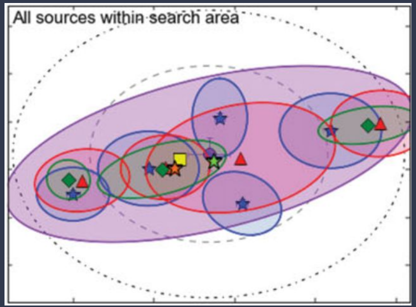


Catalogues at large $\Delta\nu$



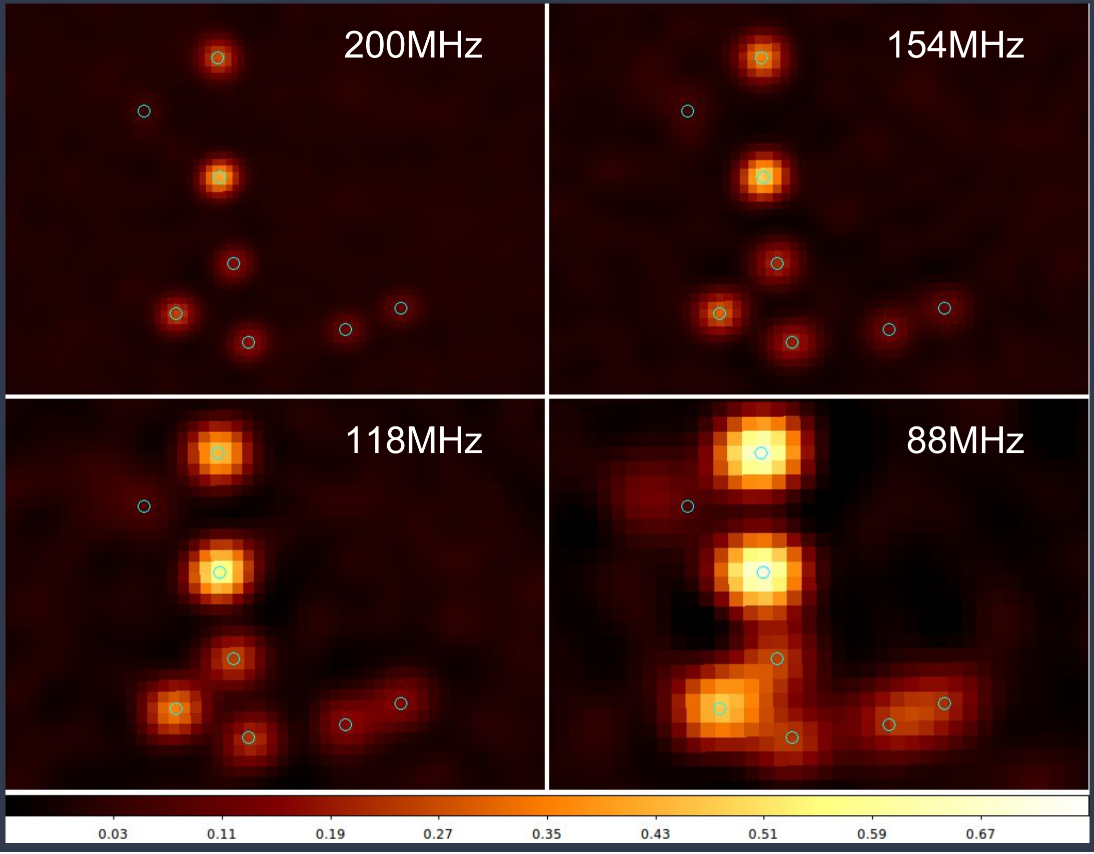
Swinbank et al. 2015

Catalogue
and
X-match?

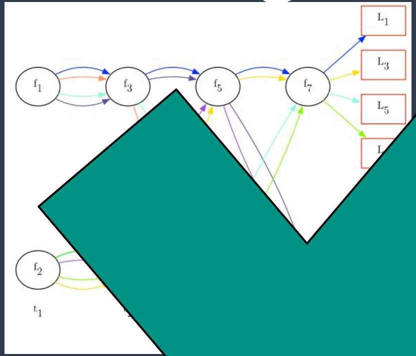


Line et al. 2017

Hierarchical
association?



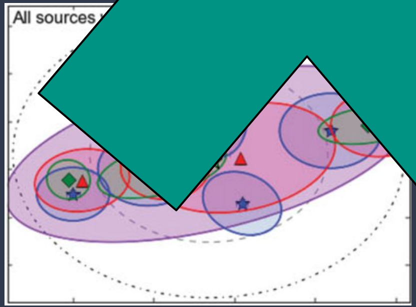
Catalogues at large $\Delta\nu$



atalogue
nd
tch?

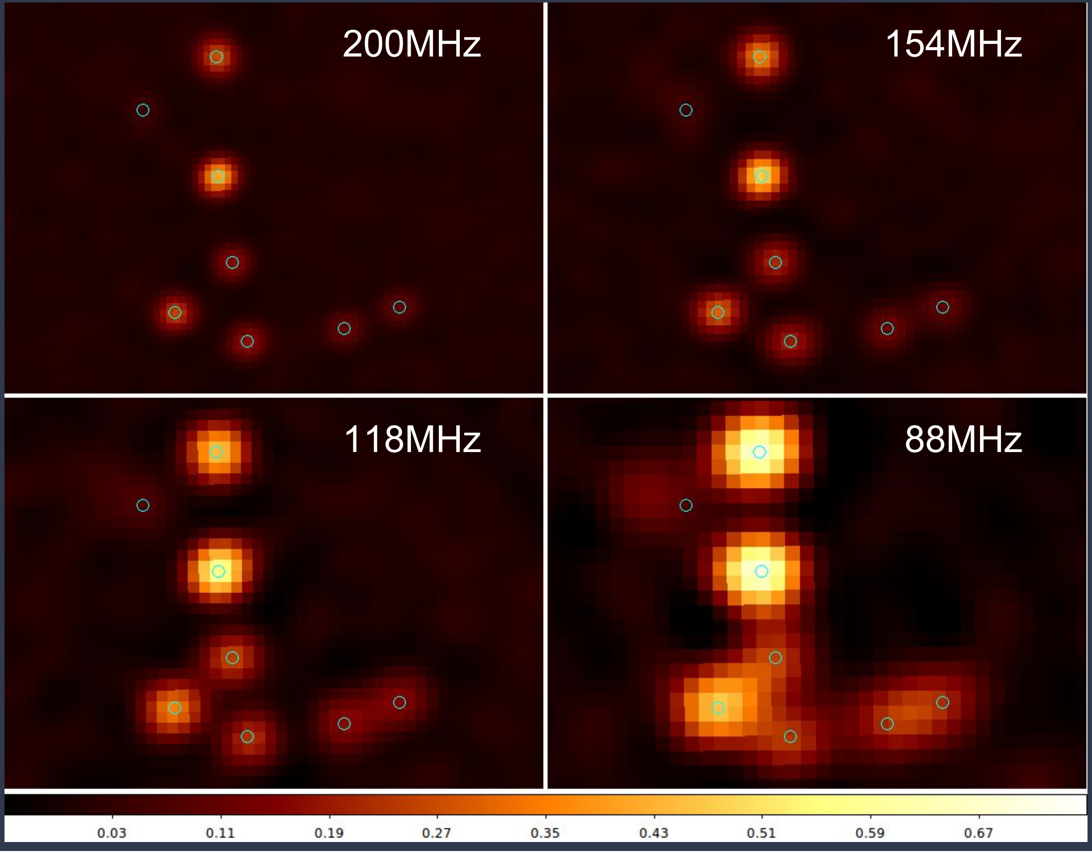
Swinbar

Too Hard!

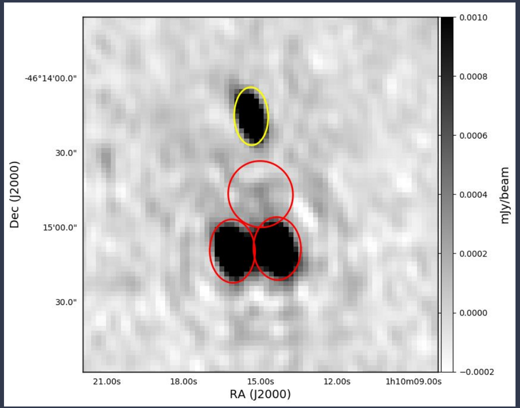


archical
ssociation?

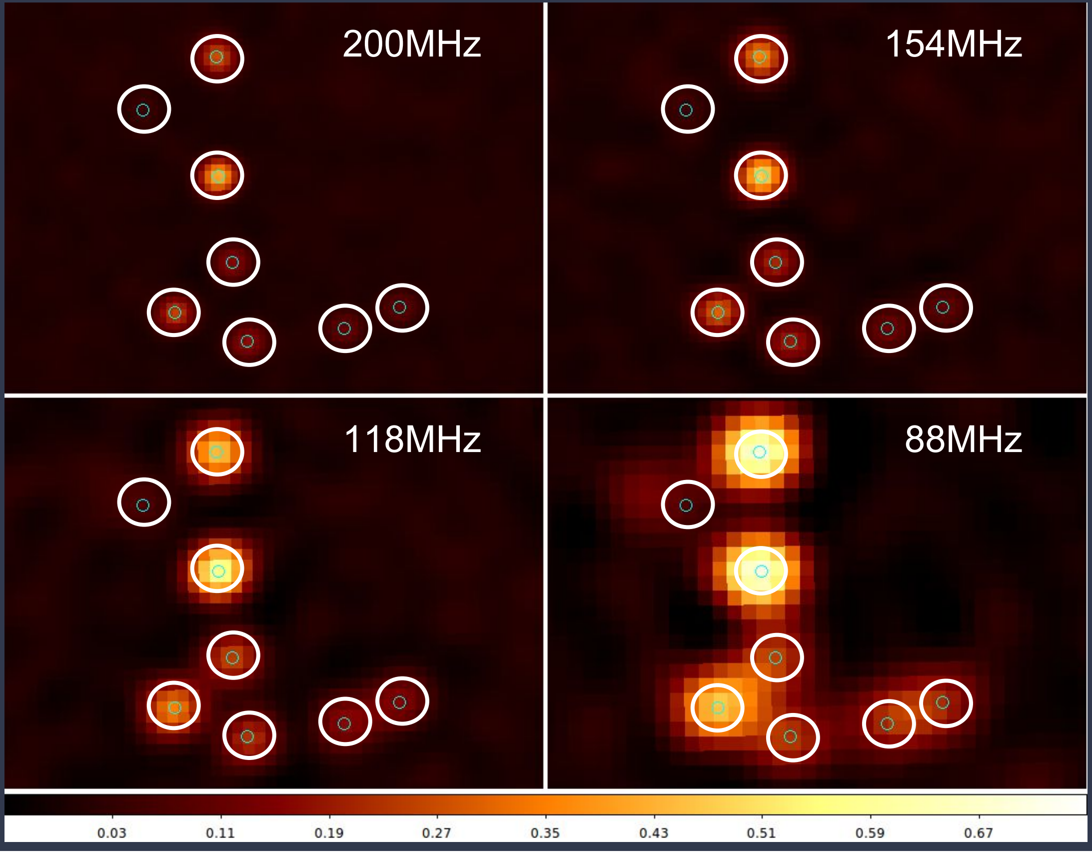
Line et al. 2017



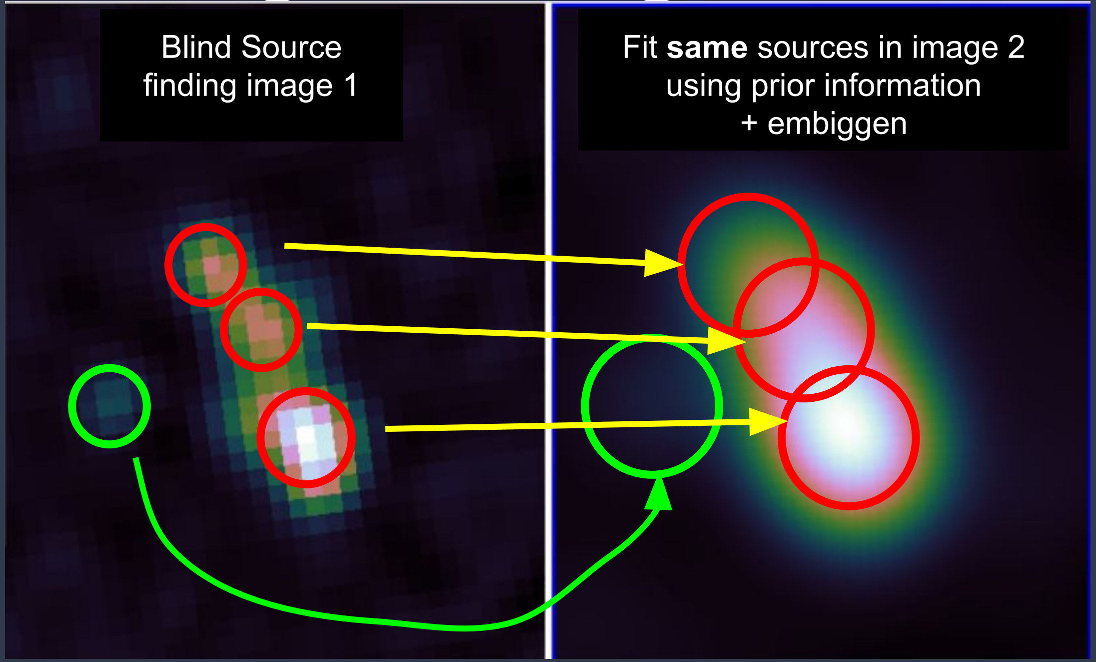
Catalogues at large $\Delta\nu$

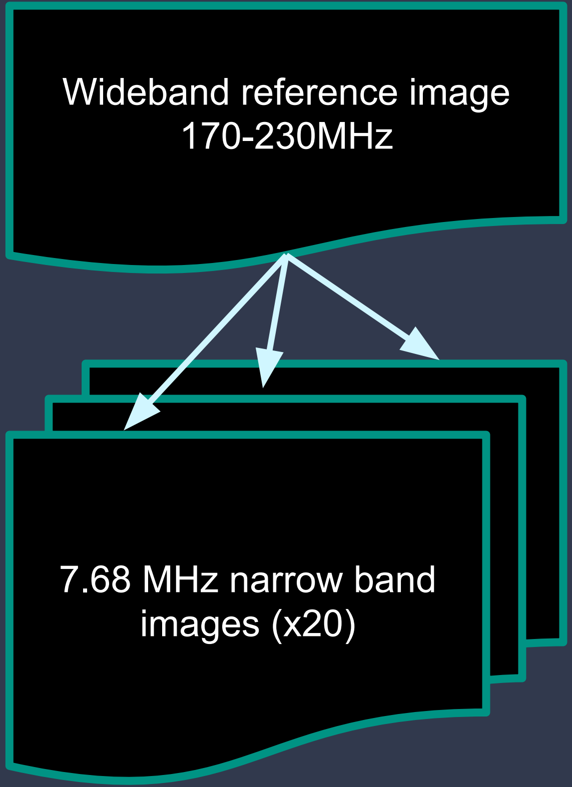


Priorized fitting with Aegean (Hancock et al. 2012/18) (now also pyBDSF)



Catalogues at large $\Delta\nu$

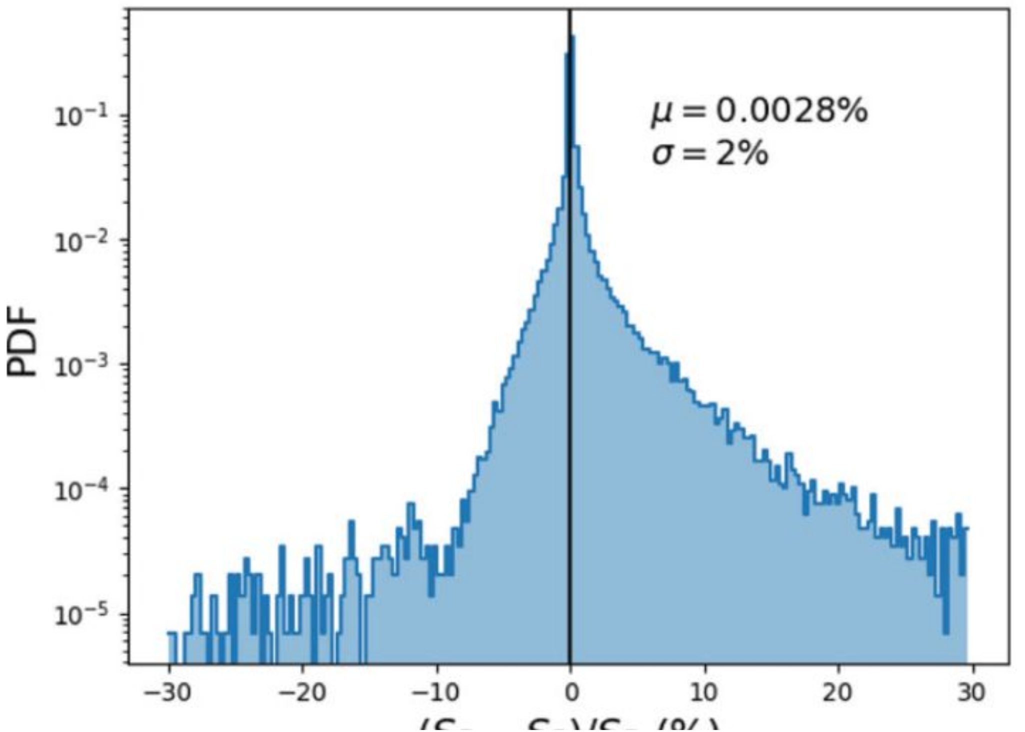




Priorized fitting

Swapping
a detection experiment for
a measurement experiment
reduces uncertainties

Good astrometry is essential so use fits_warp:
Hurley-Walker & Hancock 2018arXiv180808017H

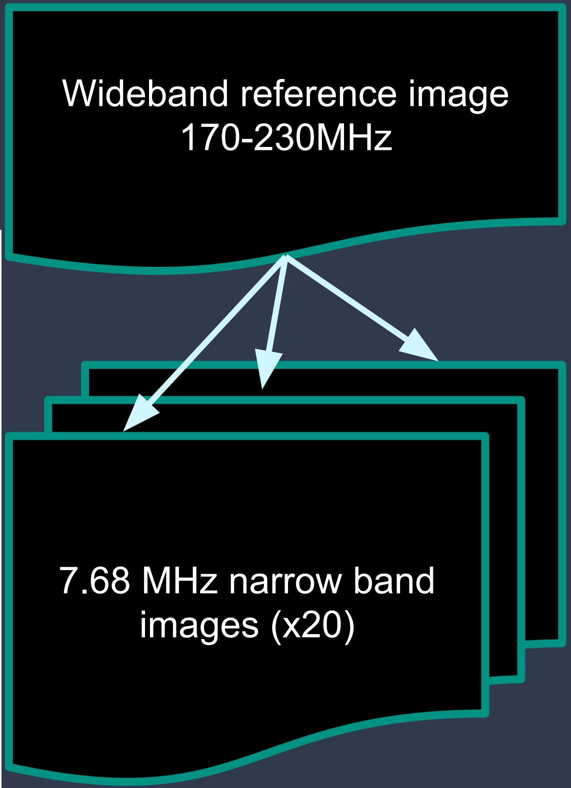


Fractional difference between Blind source detection and Priorized fitting

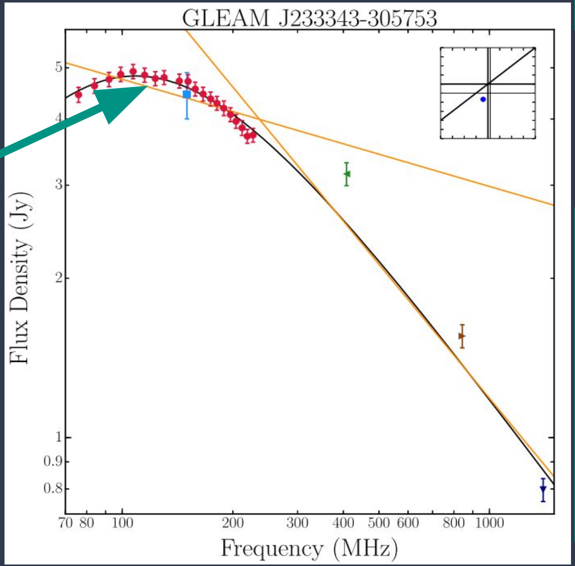
Catalogues at large $\Delta\nu$

Catalog contains

- **all** sources from deep image
- fluxes from each narrow band for each source
- sub-threshold fluxes
- ZERO false cross ids



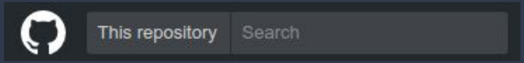
GLEAM priorized fits
at 20 frequencies



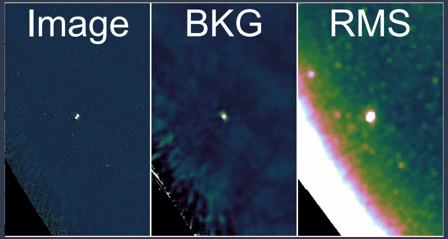
Callingham et al. 2017

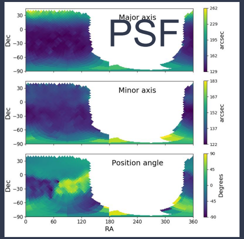
Source Finding Solution: Aegean

https://github.com/PaulHancock/Aegean



$$\text{Min}\ \{ (f(\theta;x) - x)^T C^{-1} (f(\theta;x) - x)\}$$

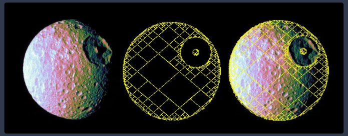




Aegean
- Find sources
- Characterise sources
- Sources can be gaussians or blobs

BANE
- Characterise background
- Characterise noise
- Do it right
- Do it fast

MIMAS
- Describe regions
- Combine regions
- Mask images
- Constrain Aegean
- Write MOC files



Other solutions:

Good ones:

- Selavy - Whiting & Humphryes 2012PASA...29..371W
- PyBDSF - Mohan & Rafferty 2015ascl.soft02007M
- PySE - Carbone et al. 2018A&C....23...92C

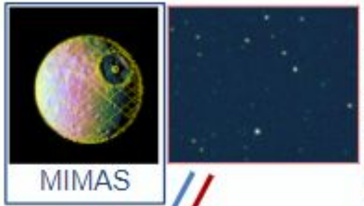
Not good ones:

- imsad (miriad)
- SAD/VSAD (aips)
- SExtractor
- Blobcat

All-in-one solutions

Survey image processing with the VAST pipeline

1. Create images. Exclude areas that are not of interest by: trimming, masking, or using MIMAS* region files. (not part of pipeline)

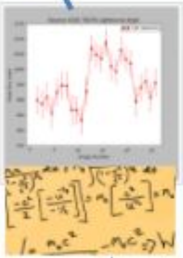


Two approaches can be used: a stream processing approach (blue flow) designed for real-time ingestion of images, and a batch processing approach (red flow) designed for already completed surveys.

7. **Publish** awesome results! (not part of pipeline, awesomeness subject to input data)

2. Create background and noise images using BANE*. **Source finding** and characterisation using Aegean*. If existing catalogues are found, then source finding is not duplicated.

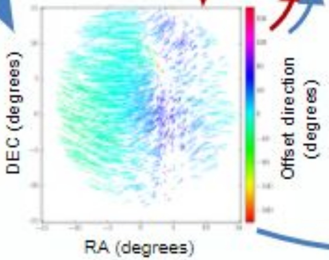
*see github.com/PaulHancock/Aegean
Hancock et al 2012, MNRAS, 422, 1812



6. **Generate light curves and calculate variability statistics** for each source. Statistics include a measure of the magnitude and confidence of variability.

5. **Flux monitoring**. For any sources which have a measurement missing from an image, replace the missing measurement with a priorized fit.

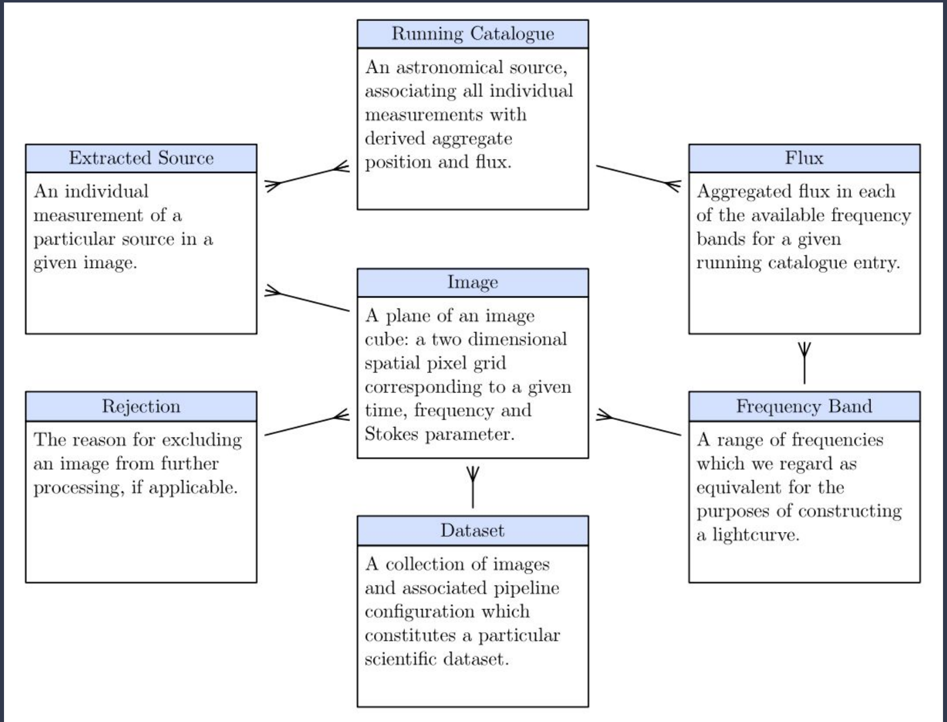
3. (optional, but recommended!) Crossmatch new measurements with a reference catalogue, and perform **astrometry and gain corrections**. This can reduce the ionospheric induced positional offsets from as much as 1arcmin, down to 5arcsec.



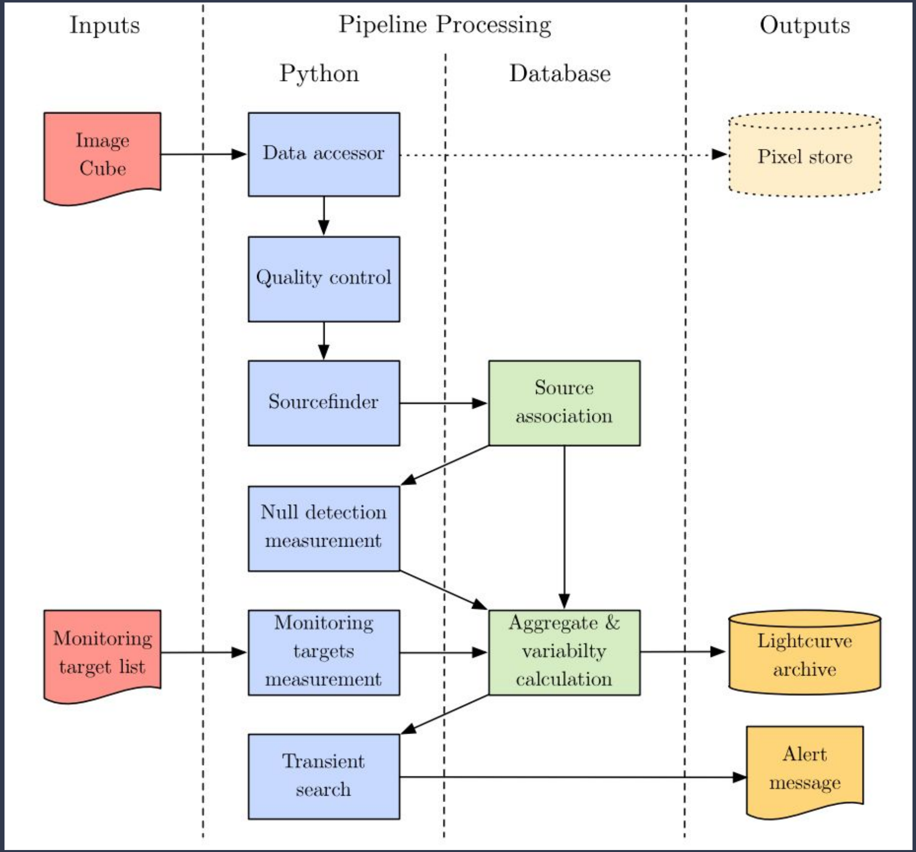
4. **Source association**: Regroup all the individual flux measurements into sources. A source will have at most one flux measurement per image.

Banyer et al 2012ASPC..461..725B

LOFAR – TraP

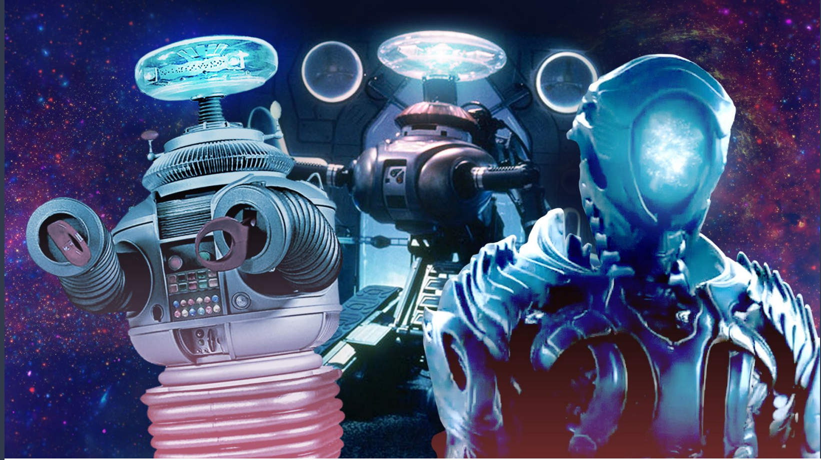


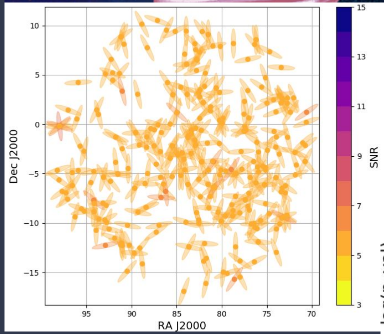
Swinbank et al. 2015A&C....11...25S

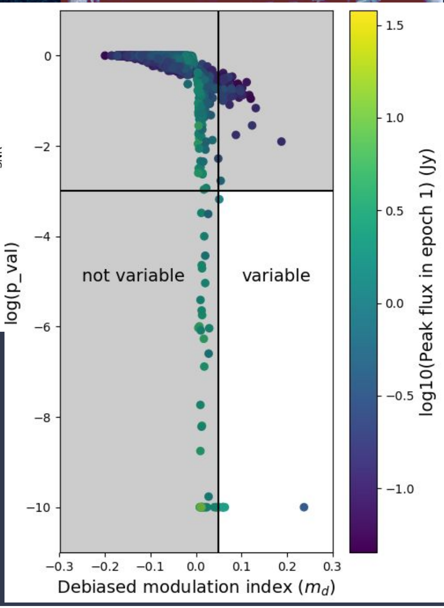


Robbie

1. Astrometry correct each epoch
2. Stack to form mean image
3. Find persistent source in mean image
4. Mask persistent sources in single epoch
5. Create light curves for persistent sources
6. Blind search for transient candidates in single epochs
7. Identify transients and characterise variability







https://github.com/PaulHancock/Robbie
(Astronomy & Computing, Submitted)

Further reading

Condon 1997PASP..109..166C Empirical measure of errors
Refreiger & Brown 1998 arXiv:9803279 analytical treatment of uncertainty and bias
Hancock et al. 2012 2012MNRAS.422.1812H Source finding with Aegean
Hancock et al. 2018 2018PASA...35...11H Source finding on correlated data
Whiting & Humphryes 2012PASA...29..371W ASKAP soft source finder
Mohan & Rafferty 2015ascl.soft02007M LOFAR source finder
PySE - Carbone et al. 2018A&C....23...92C LOFAR source finder (for TraP)
Hurley-Walker & Hancock 2018 2018arXiv180808017H Correcting ionospheric effects in the image plane
Banyer et al 2012ASPC..461..725B VAST pipeline
Hancock et al. 2018ascl.soft08011H Robbie (= vast lite / vast ++)
Swinbank et al. 2015A&C....11...25S LOFAR TraP