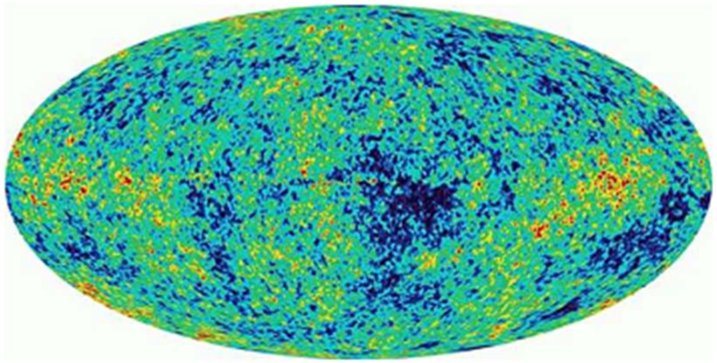
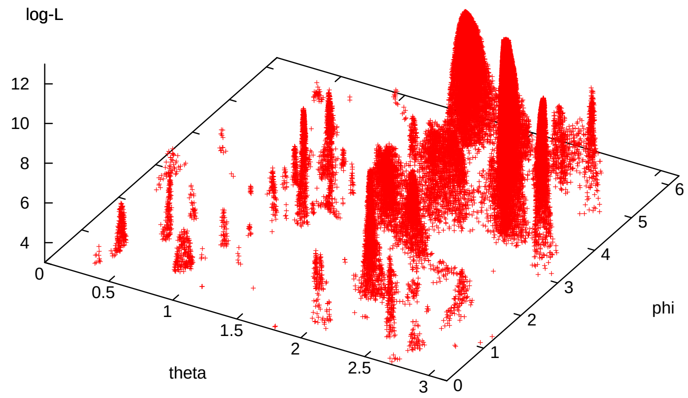
Bayesian Source Detection





Farhan Feroz
f.feroz@mrao.cam.ac.uk

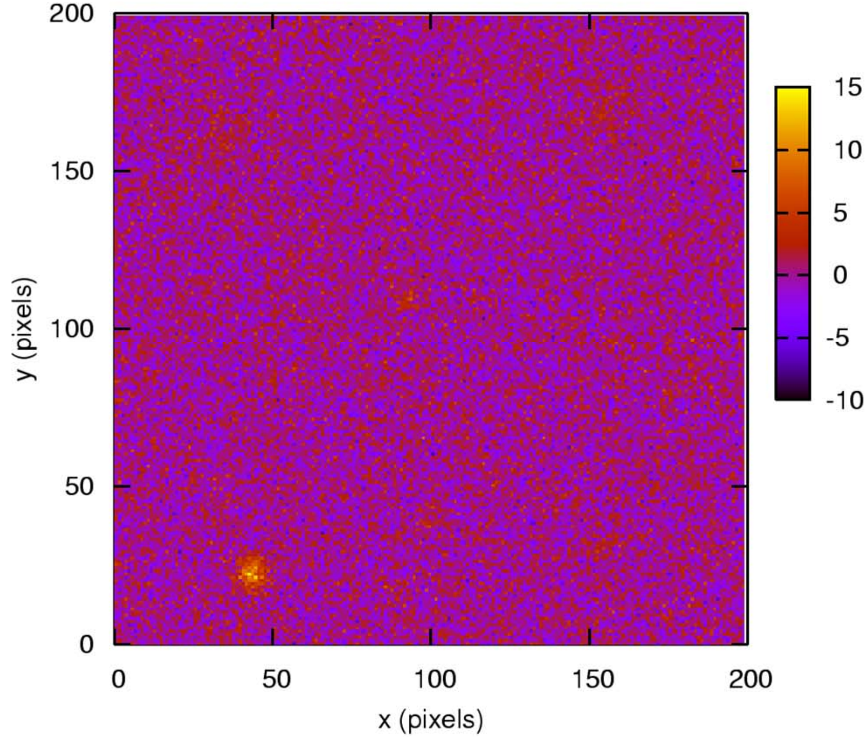
Cavendish Astrophysics

UNIVERSITY OF CAMBRIDGE

Source Detection Problems in Astrophysics

Feroz & Hobson (2008, MNRAS, 384, 449)

How many sources?



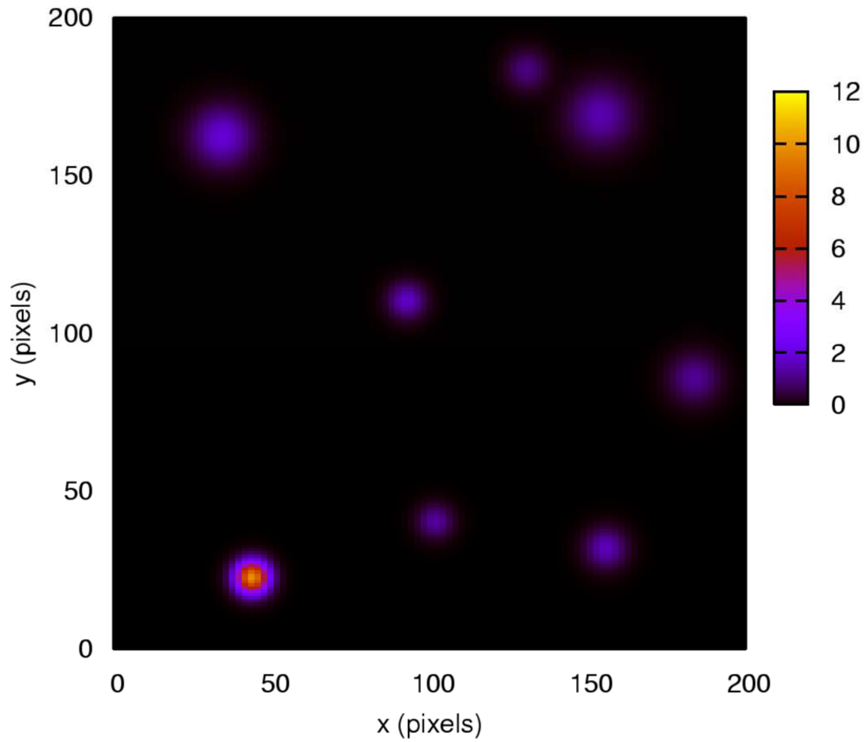
Signal + Noise

SNR ~ 0.5 – 1.0

Source Detection Problems in Astrophysics

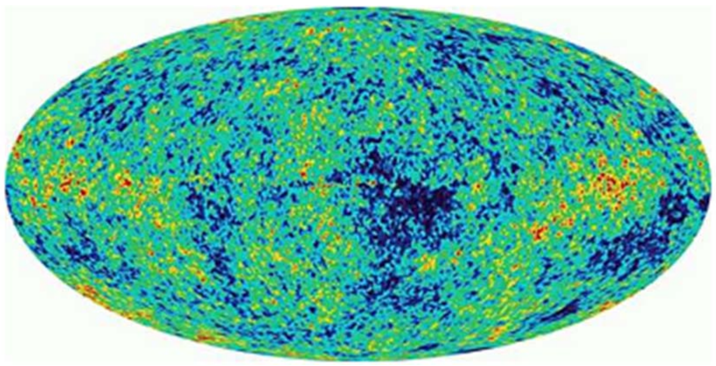
Feroz & Hobson (2008, MNRAS, 384, 449)

How many sources?

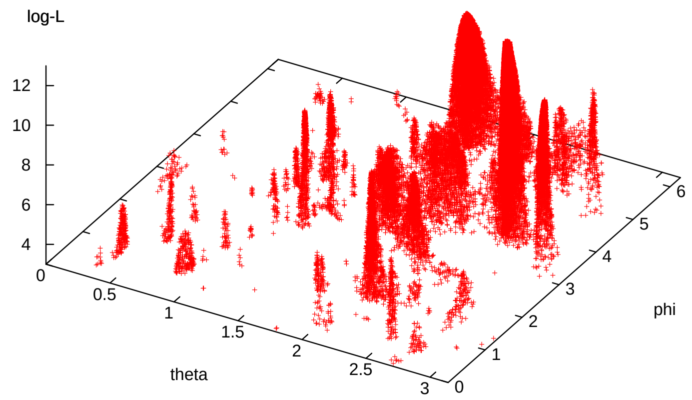


Signal (8 sources)

Probabilistic Source/Object Detection



- Problems in Source Detection
  - Identification
  - Quantifying Detection
  - Characterization



Textures in CMB

Bayesian Parameter Estimation

- Collect a set of N data points D_i ($i = 1, 2, \ldots, N$), denoted collectively as data vector **D**.

- Propose some model (or hypothesis) H for the data, depending on a set of M parameter θ_i ($i = 1, 2, \ldots, N$), denoted collectively as parameter vector **θ**.

- Bayes' Theorem:

Likelihood Prior

$$\mathrm{P}(\boldsymbol{\theta} \mid \mathbf{D}, H) = \frac{\mathrm{P}(\mathbf{D} \mid \boldsymbol{\theta}, H)\,\mathrm{P}(\boldsymbol{\theta} \mid H)}{\mathrm{P}(\mathbf{D} \mid \mathrm{H})} \longrightarrow \mathrm{P}(\boldsymbol{\theta}) = \frac{\mathrm{L}(\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\mathrm{Z}}$$

Posterior Evidence

- Parameter Estimation: $\mathrm{P}(\boldsymbol{\theta})\ \alpha\ \mathrm{L}(\boldsymbol{\theta})\pi(\boldsymbol{\theta})$

  posterior α likelihood x prior

Bayesian Model Selection

$$P(\boldsymbol{\theta} \mid \mathbf{D}, H) = \frac{P(\mathbf{D} \mid \boldsymbol{\theta}, H) P(\boldsymbol{\theta} \mid H)}{P(\mathbf{D} \mid \mathrm{H})} \rightarrow P(H \mid \mathbf{D}) = \frac{P(\mathbf{D} \mid H) P(H)}{P(\mathbf{D})}$$

- Consider two models H_0 and H_1

$$R = \frac{P(H_1 \mid \mathbf{D})}{P(H_0 \mid \mathbf{D})} = \frac{P(\mathbf{D} \mid H_1) P(H_1)}{P(\mathbf{D} \mid H_0) P(H_0)} = \frac{Z_1}{Z_0} \frac{P(H_1)}{P(H_0)}$$

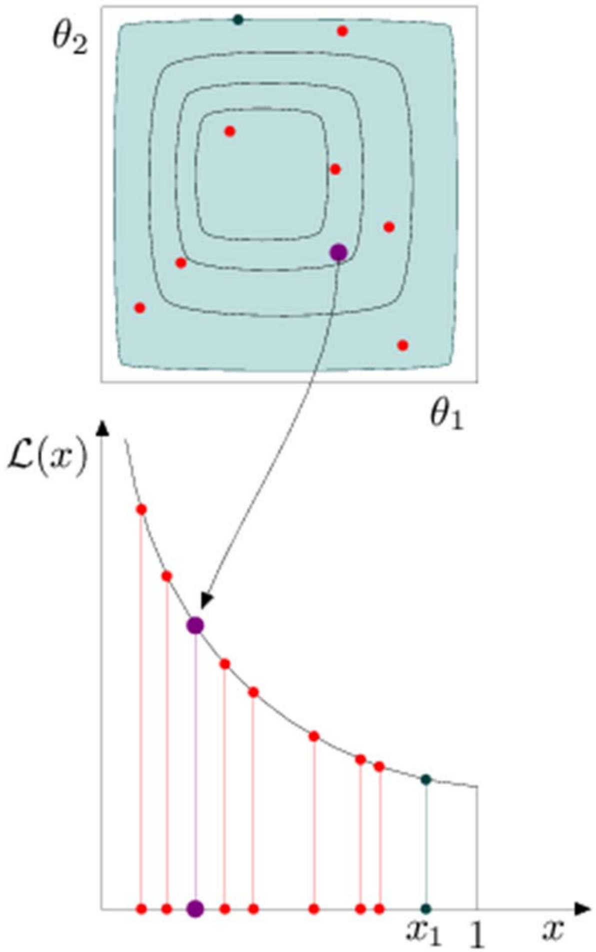
Bayes Factor

Prior Odds

- Bayesian Evidence $Z = P(\mathbf{D}|H) = \int L(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}$ plays the central role in Bayesian Model Selection.

- Bayesian Evidence rewards model preditiveness.
  - Sets more stringent conditions for the inclusion of new parameters

Nested Sampling: Algorithm



1. Sample N 'live' points uniformly inside the initial prior space ($X_0 = 1$) and calculate their likelihoods

2. Find the point with the lowest L_i and remove it from the list of 'live' points

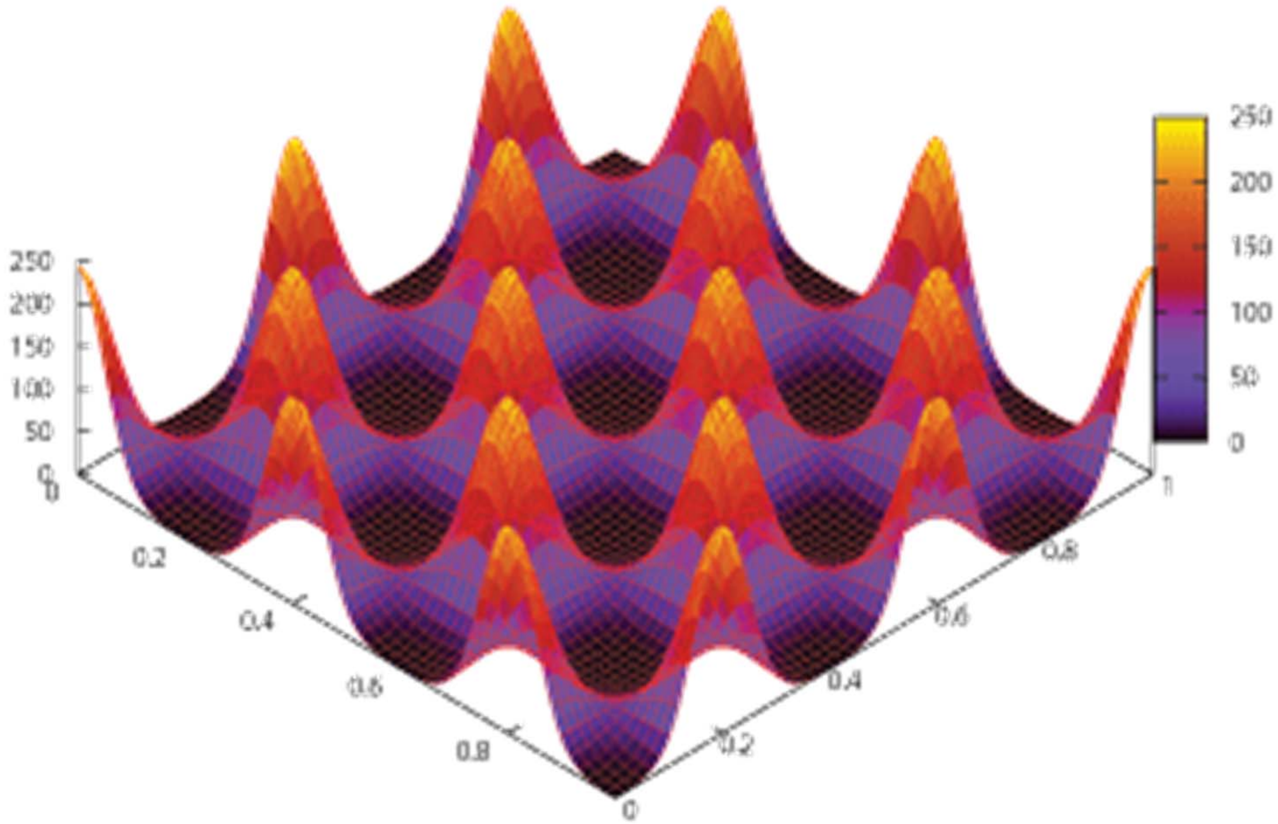
3. Increment the evidence as $Z = Z + L_i \, (X_{i-1} - X_{i+1}) \, / \, 2$

4. Reduce the prior volume $X_i \, / \, X_{i-1} = t_i$ where $P(t) = N \, t^{N-1}$

5. Replace the rejected point with a new point sampled from $\pi(\theta)$ with constraint $L > L_i$

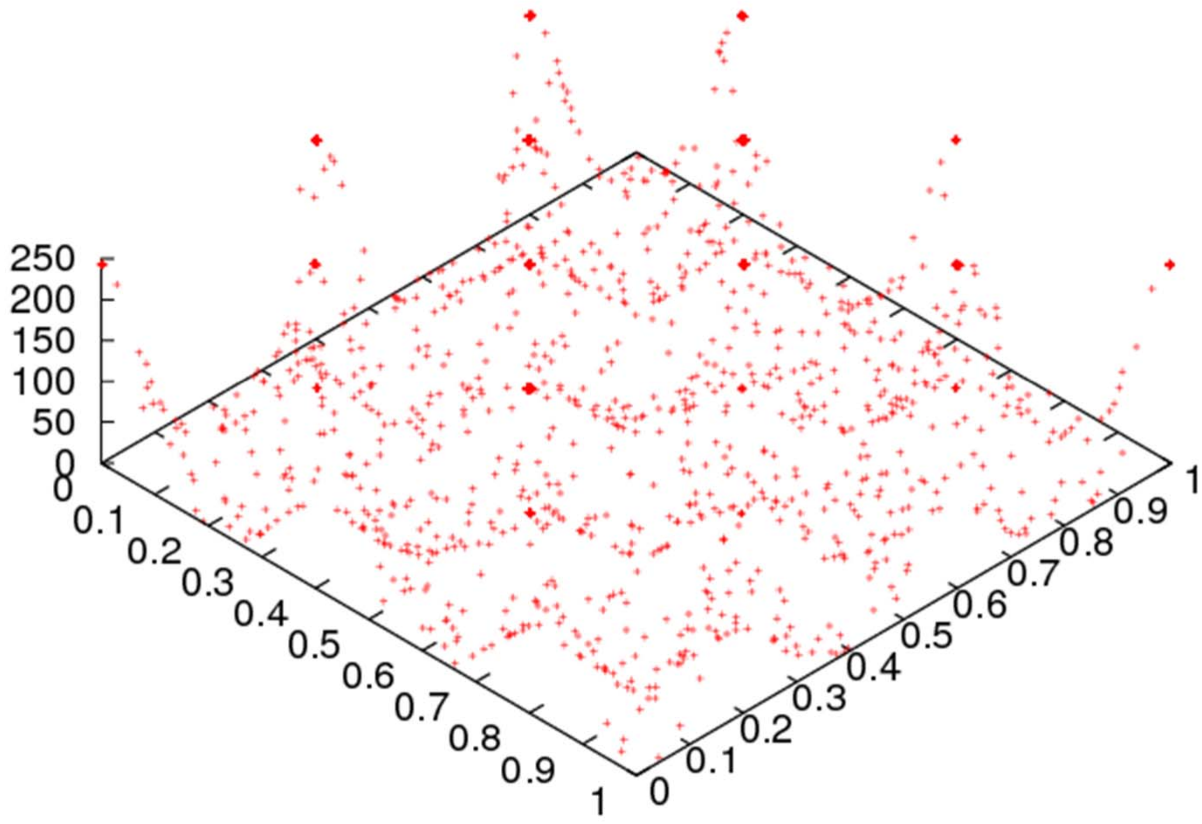
6. If $L_{\max} X_i < \alpha Z$ then stop else goto 3

Nested Sampling: Demonstration



Egg-Box Posterior

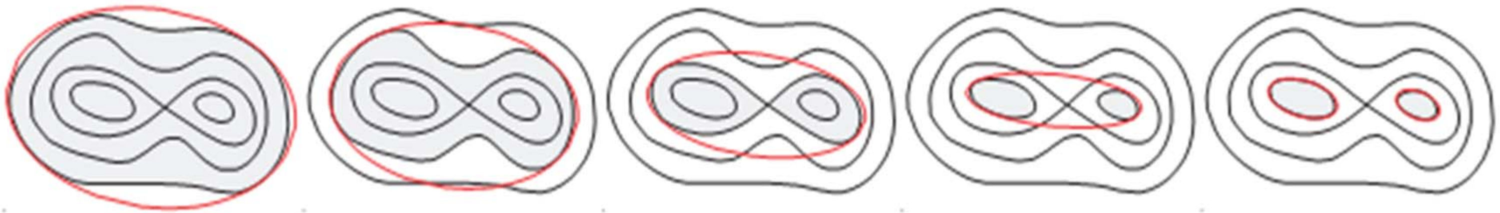
Nested Sampling: Demonstration



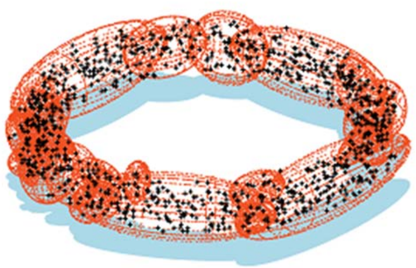
Egg-Box Posterior

Multi-modal Nested Sampling (MultiNest)

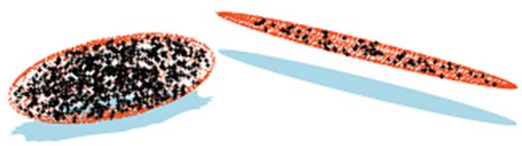
- Introduced by Feroz & Hobson (2008, MNRAS, 384, 449, arXiv:0704.3704), refined by Feroz, Hobson & Bridges (2009, MNRAS, 398, 1601, arXiv:0809.3437)



Ellipsoidal Rejection Sampling



Uni-Modal Distribution



Multi-Modal Distribution

Probabilistic Models for Source Detection

Bayesian Source Detection: Variable Source Number Model

- Bayesian Purist Gold Standard: detect and characterize all sources in the data simultaneously ⇒ infer full parameter set $\boldsymbol{\theta} = \{N_s, p_1, p_2, \ldots, p_{N_s}, q\}$ where
  N_s = number of sources
  p_i = parameters associated with i[th] source
  q = parameters common to all the sources

- Allows straight-forward inclusion of prior information on number of sources, N_s.

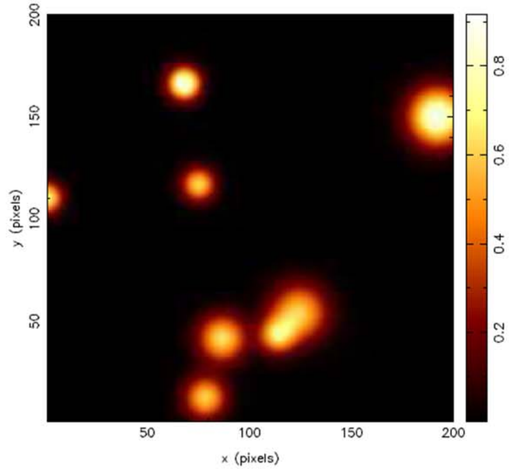
- Complication
  - Length of parameter vector, **θ**, is variable
  - Requires reversible-jump MCMC (see Green, 1995, *Biometrika*, V. 82)
  - Counting degeneracy when assigning source parameters in each sample to sources in image ⇒ at least $N_s!$ modes

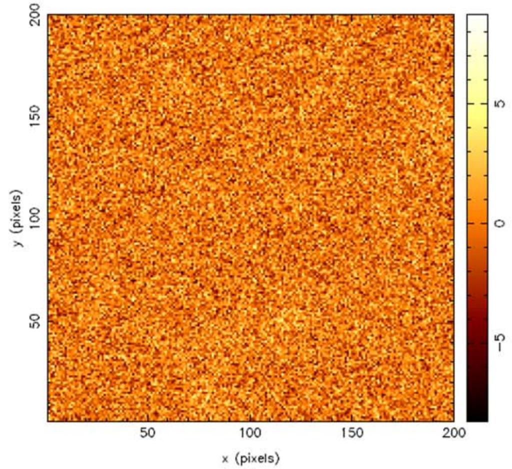
- Practical Concern: If prior on N_s remains non-zero at large N_s
  - Parameter space to be explored becomes very large
  - Slow mixing, can be very inefficient

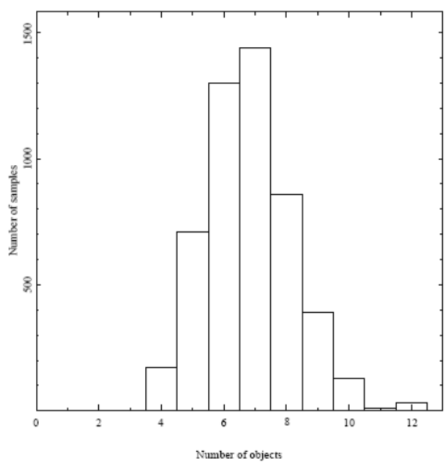
Bayesian Source Detection: Variable Source Number Model

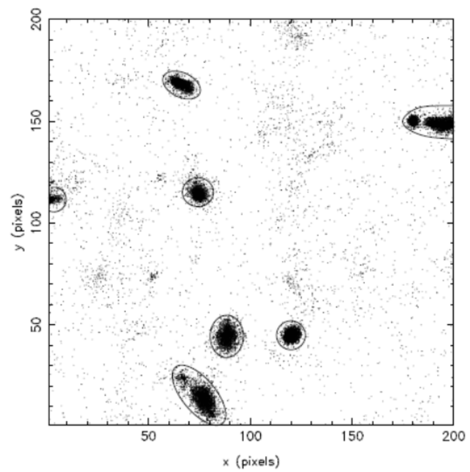
Hobson & McLachlan, 2002, astro-ph/0204457

- 8 Gaussian sources, with variable scale and amplitude, in Gaussian noise
- Analysis done with BayeSys (http://www.inference.phy.cam.ac.uk/bayesys/)
  - Runtime: 17 hours CPU time









Bayesian Source Detection: Fixed Source Number Model

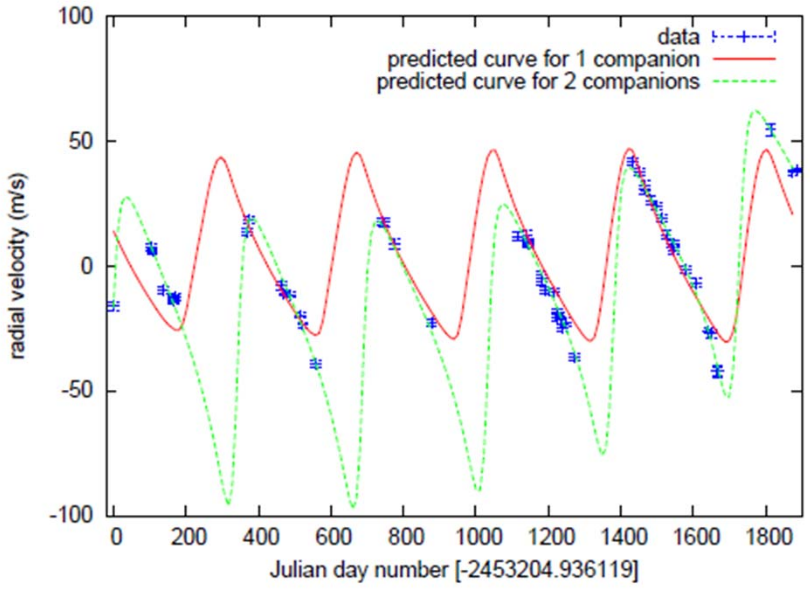
- Poor man's approach to Bayesian gold standard

- Consider series of models H_{N_s}, each with fixed N_s, where N_s goes from say 0 to N_{max}
  - Length of parameter space is fixed for each model
  - Can use standard MCMC or nested sampling

- Determine preferred number of source using Bayesian model selection

Applications: Exoplanet Detection – HIP 5158

Feroz, Balan & Hobson, 2011, arXiv:1105.1150



N_p	$\ln \mathcal{Z}$	s (m/s)
1	30.05 ± 0.14	10.31 ± 1.09
2	78.19 ± 0.15	2.28 ± 0.31
3	70.68 ± 0.16	2.28 ± 0.28

Table 2. The evidence and jitter values for the system HIP 5158. The null evidence (-255.3) has been subtracted from each $\ln \mathcal{Z}$ value.

Bayesian Source Detection: Single Source Model

- Special case of fixed source number model, simply set $N_s = 1$

- Not restricted to detecting just one source in the data
  - Trade-off high dimensionality with multi-modality
  - Posterior will have numerous modes
  - Each corresponding to a either real or spurious source

- Fast and reliable method when sources (effects) are non-overlapping

- Use local evidences for distinguishing between real and spurious sources

Quantifying Source Detection: Single Source Model

- $$p_{TP} = \frac{R}{1+R}$$

- $$R = \frac{\Pr(H_1 \mid \mathrm{D})}{\Pr(H_0 \mid \mathrm{D})} = \frac{\Pr(\mathrm{D} \mid H_1)\Pr(H_1)}{\Pr(\mathrm{D} \mid H_0)\Pr(H_0)} = \frac{Z_1 \Pr(H_1)}{Z_0 \Pr(H_0)}$$

- H_0 = "there is no source with its centre lying in the region S"
- H_1 = "there is one source with its centre lying in the region S"

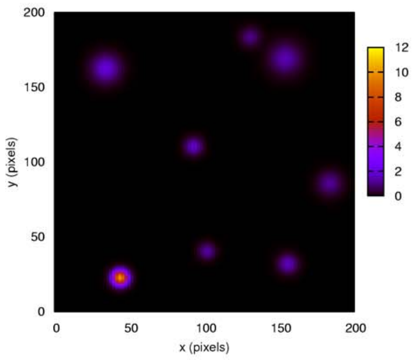
- $$Z_0 = \frac{1}{|S|}\int_s L_0 dX = L_0$$

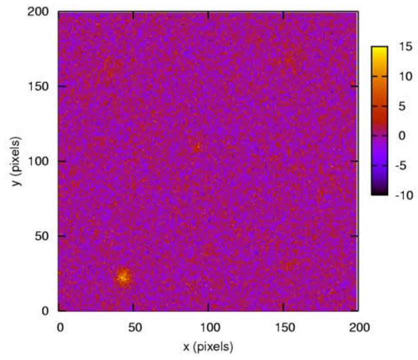
- For sources distributed according to Poisson distribution

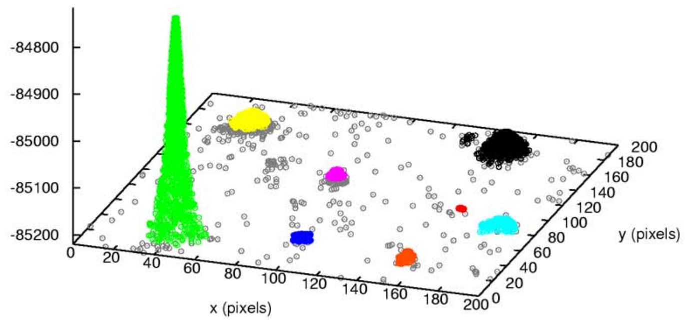
$$\frac{\Pr(H_1)}{\Pr(H_0)} = \mu_s, \quad \therefore R = \frac{Z_1 \mu_s}{L_0}$$

How Many Sources? Bayesian Solution

Feroz & Hobson, 2008, MNRAS, 384, 449, arXiv:0704.3704







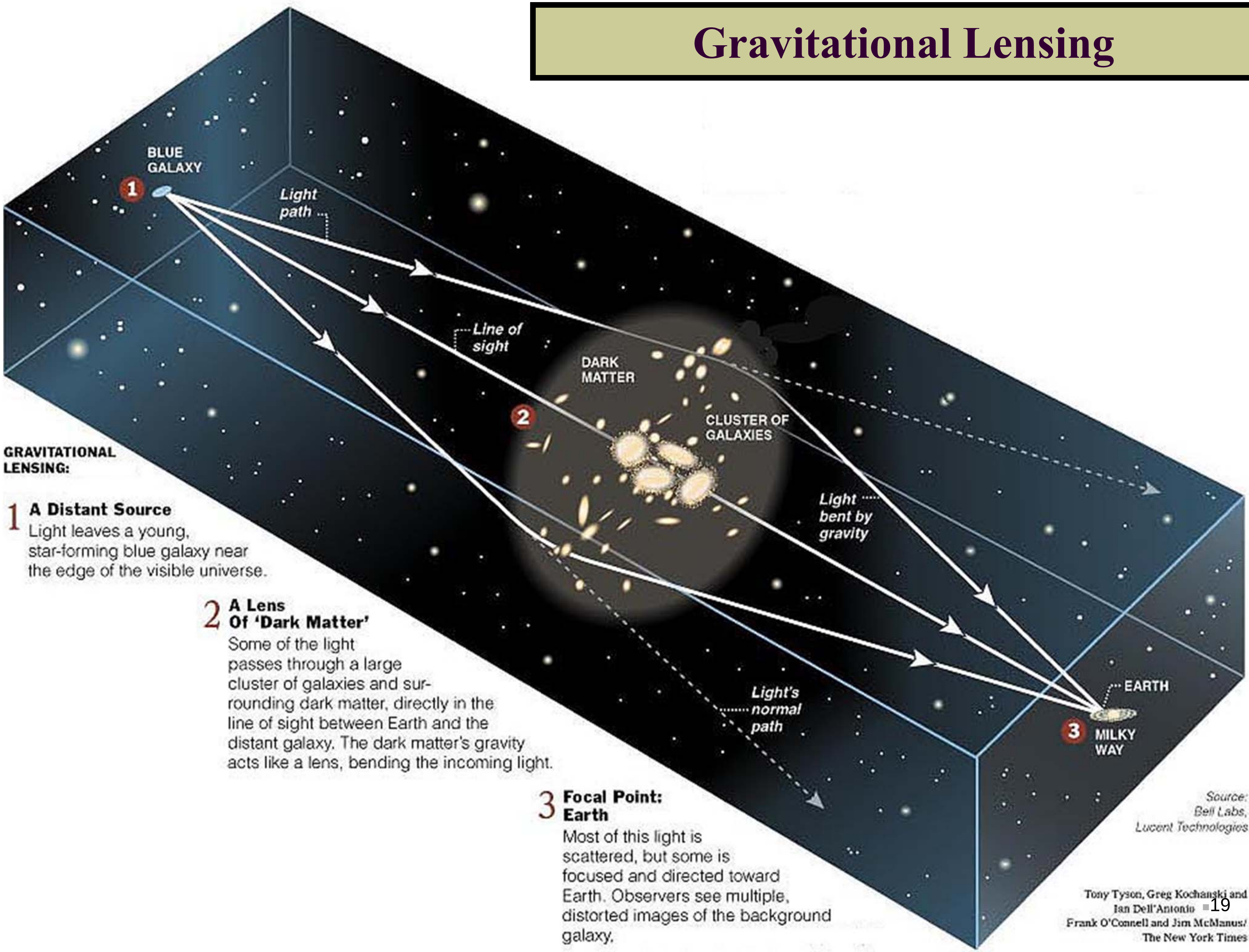
MultiNest

- 7 out of 8 objects identified
  - missed 1 object because 2 objects are very close
- *runtime* = 2 *min* on a normal desktop

Thermodynamic Integration

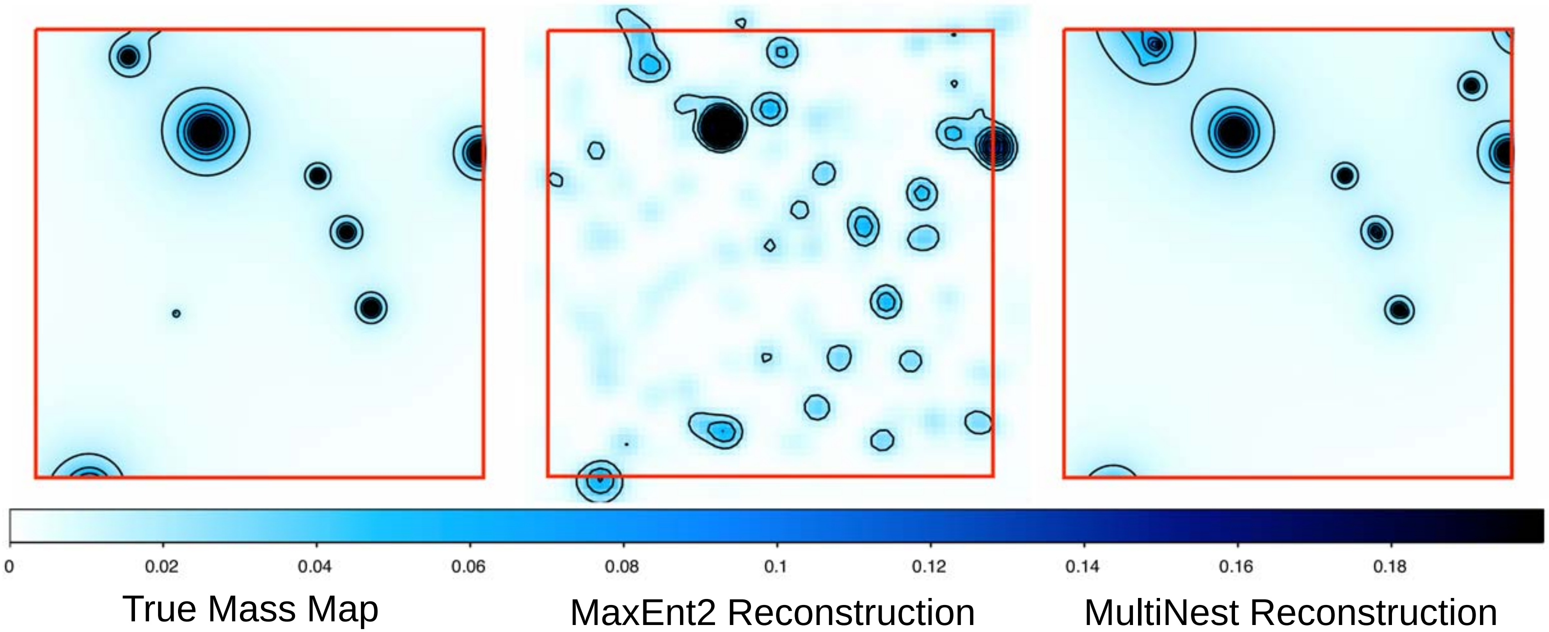
- Solution possible only through iterative sampling (see McLachlan & Hobson, 2002)
- *runtime* > 16 *hours* on a normal desktop

Gravitational Lensing



Applications: Clusters in Weak Lensing

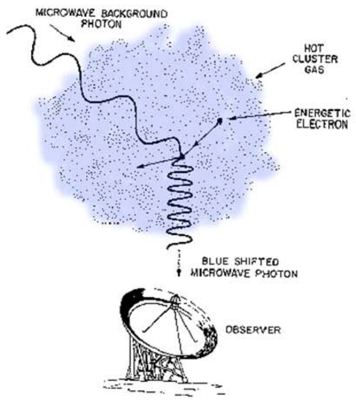
Feroz, Marshall & Hobson, 2008, arXiv:0810.0781

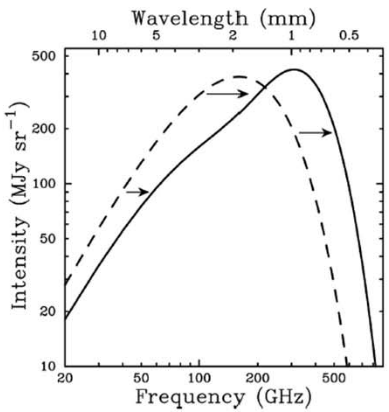


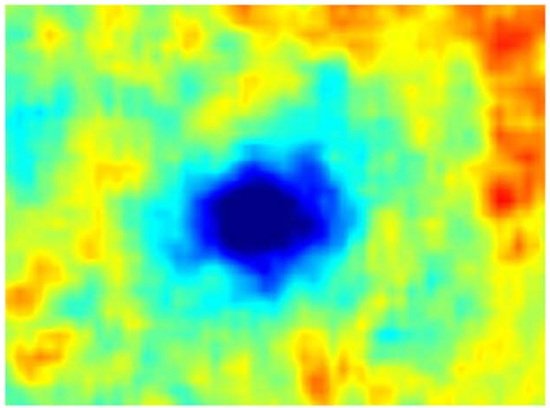
- 0.5 x 0.5 degree^2, 100 gal per arcmin^2 & $\sigma = 0.3$
- Concordance ΛCDM Cosmology with cluster mass & redshifts drawn from Press-Schechter mass function
- $p_{\mathrm{th}} = 0.5$

Clusters in Sunyaev-Zel'dovich (SZ)

- Distortion of CMB by hot intra-cluster electrons through inverse Compton scattering



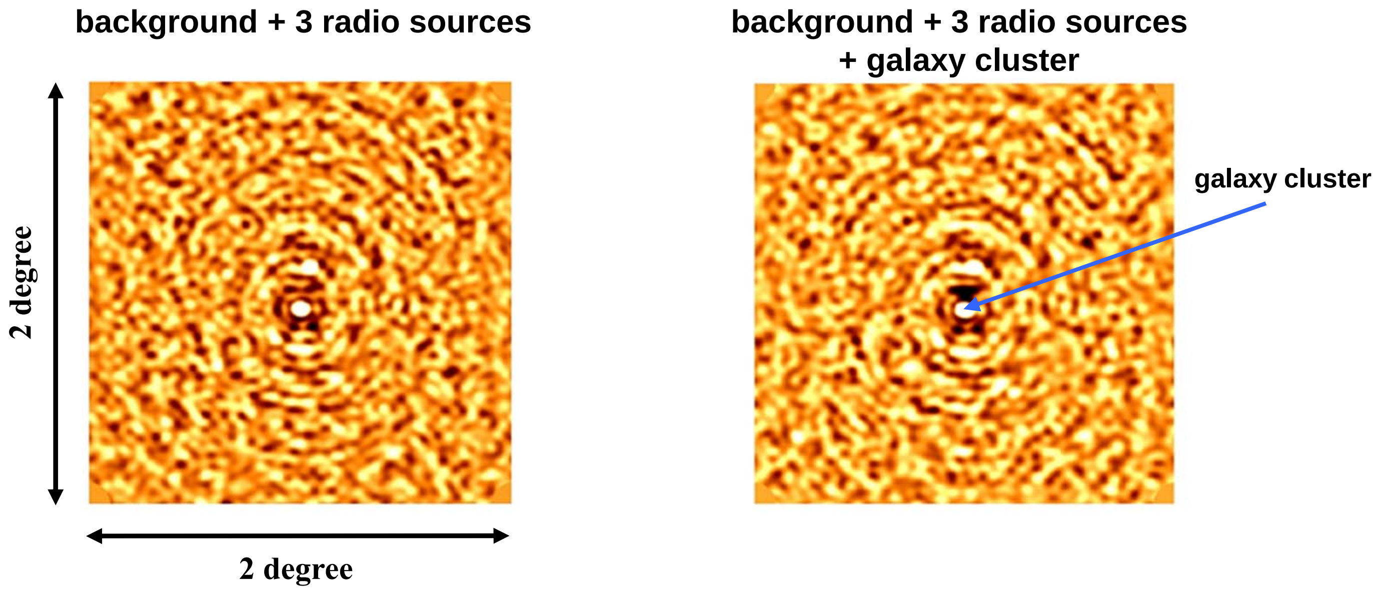




Applications: Clusters in Sunyaev Zel'dovich (SZ)

Feroz et al., 2009, arXiv:0811.1199

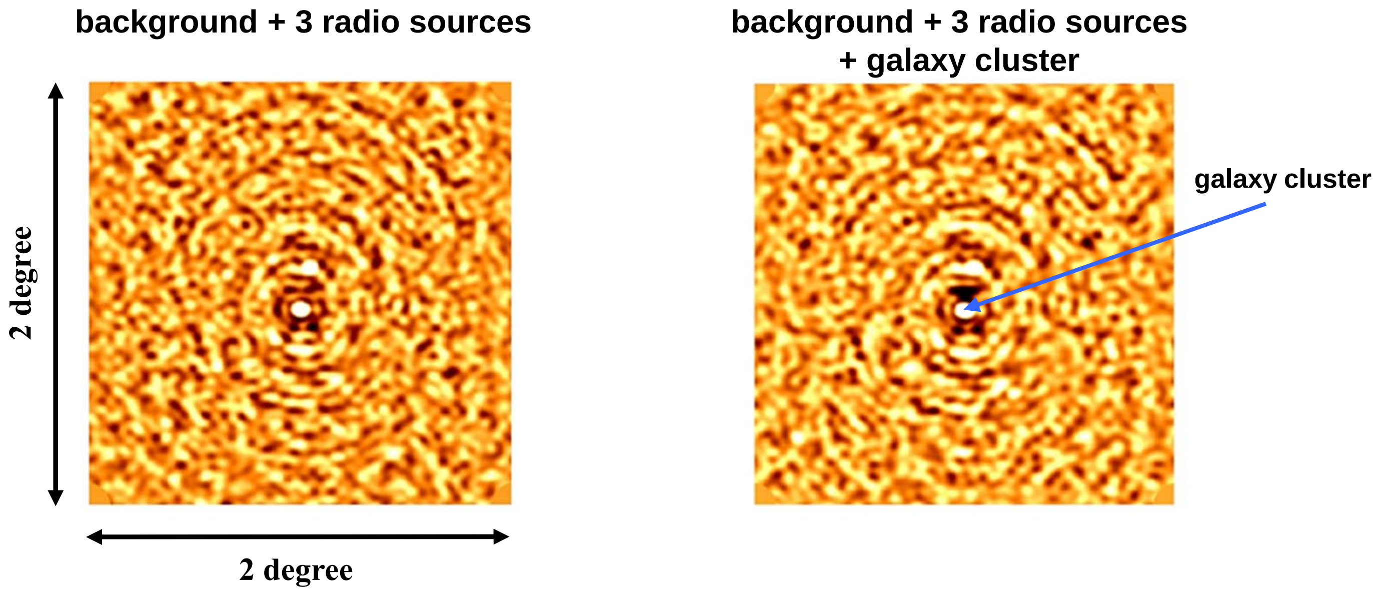
Galaxy cluster (and radio sources) in interferometric SZ data



Applications: Clusters in Sunyaev Zel'dovich (SZ)

Feroz et al., 2009, arXiv:0811.1199

Galaxy cluster (and radio sources) in interferometric SZ data



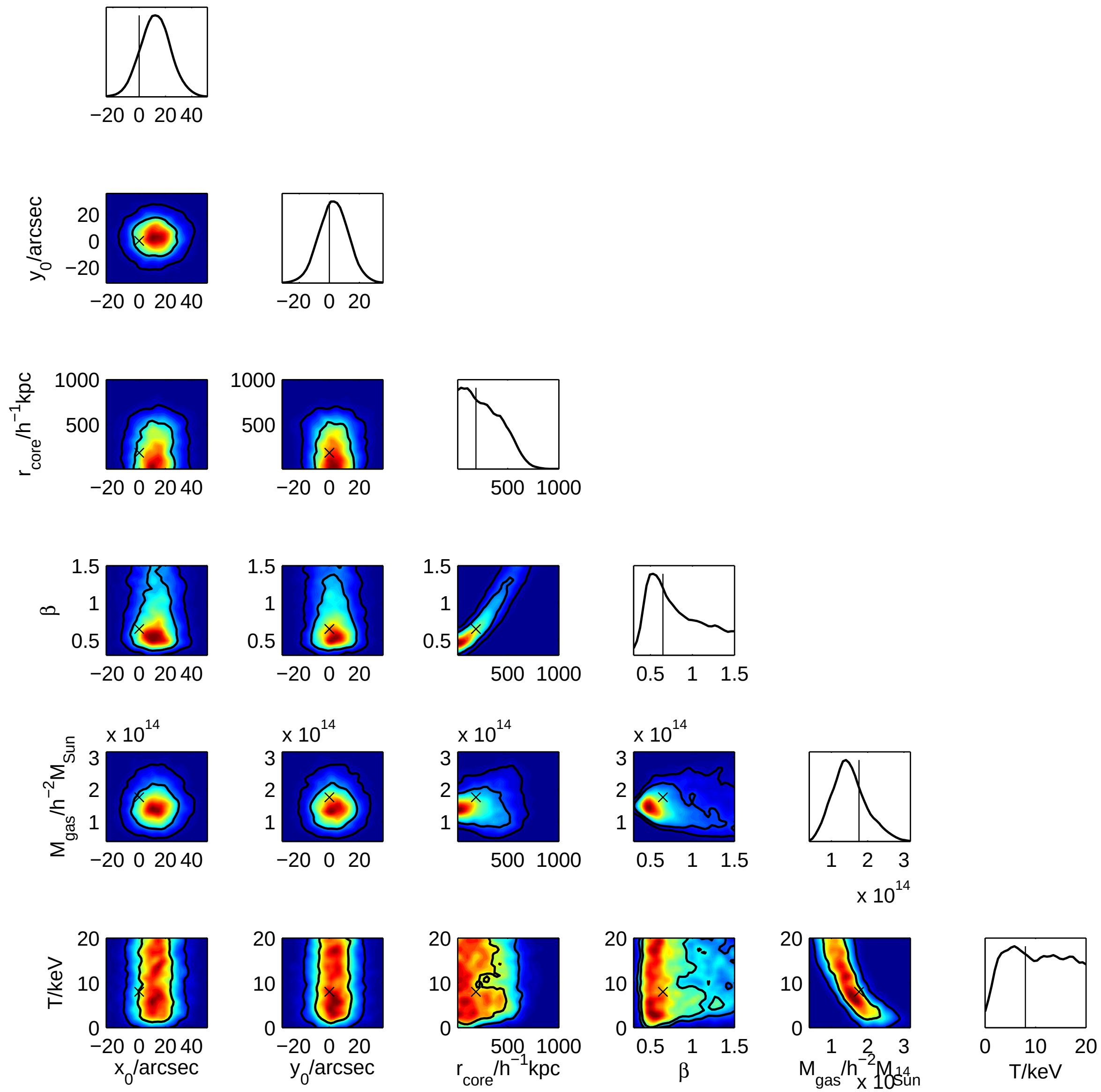
Bayesian Model Comparison

R = P(cluster | data)/P(no cluster | data)

R = 0.35   **R ~ 10^{33}**

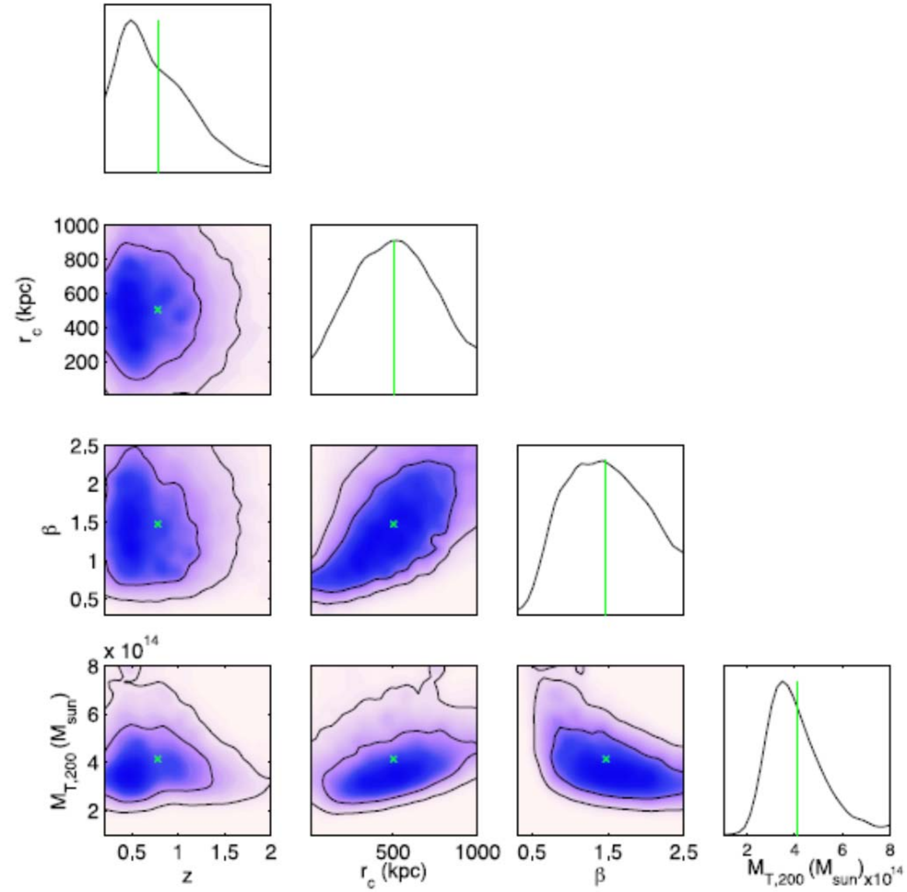
Clusters in SZ – Parameter Constraints

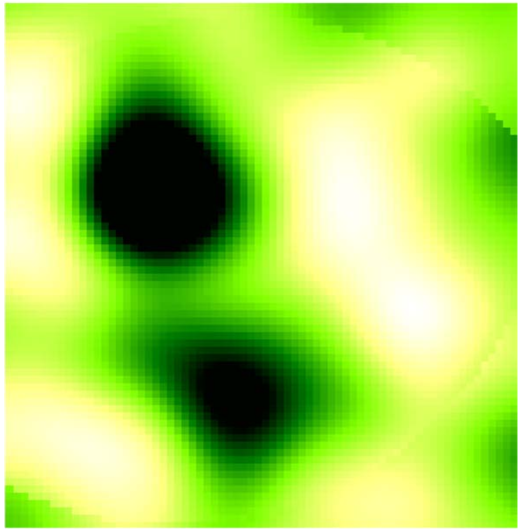
Feroz et al., 2009, arXiv:0811.1199



First AMI Blind Cluster: AMI-CL J0300+2613

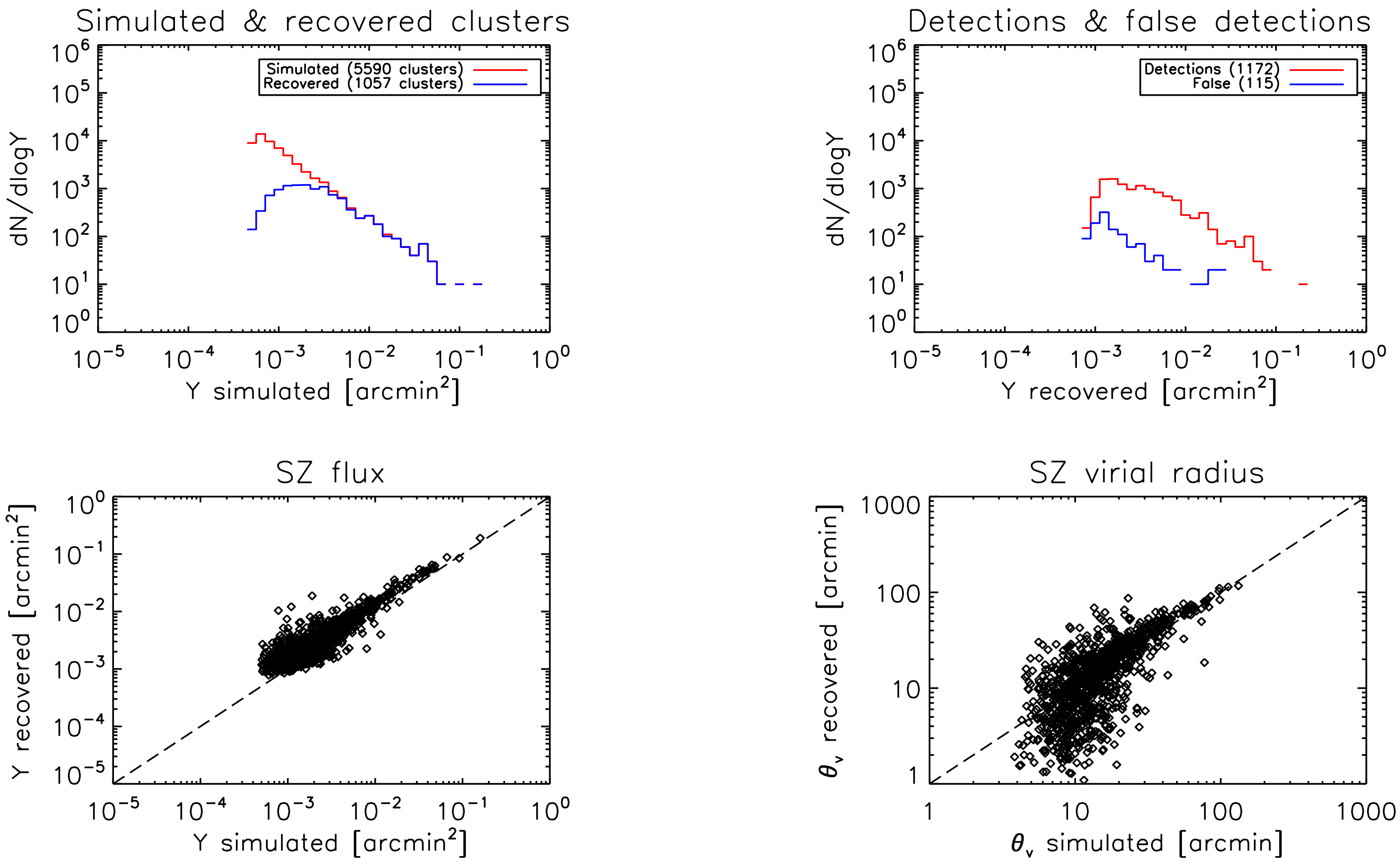
arXiv:1012.4441, 1305.6655





Data	$M_{T,200}$ $\times 10^{14}\,M_\odot$	z	β	r_c (kpc)	$T_{e,200}$ (keV)	Blind $\ln\left(\frac{Z_1(S)}{Z_0}\right)$	R	Known $\ln\left(\frac{Z_1(S)}{Z_0}\right)$
AMI	$5.1^{+1.1}_{-1.2}$	$0.69^{+0.31}_{-0.30}$	$1.6^{+0.5}_{-0.5}$	440^{+220}_{-220}	$5.3^{+1.1}_{-1.1}$	9.49	4.5×10^3	–
AMI and SZA	$4.1^{+1.1}_{-1.1}$	$0.78^{+0.40}_{-0.38}$	$1.5^{+0.6}_{-0.6}$	500^{+260}_{-250}	$4.7^{+1.1}_{-1.1}$	11.06	5.4×10^3	–
SZA	$2.2^{+0.4}_{-0.5}$	$0.9^{+0.36}_{-0.35}$	$1.6^{+0.5}_{-0.5}$	460^{+250}_{-250}	$3.3^{+0.6}_{-0.6}$	1.24	0.29	2.55

Planck SZ Challenge II – Results with MultiNest



- 50 x 10^6 pixels, ~ 1000 recovered clusters, ~ 3 CPU hours

Bayesian Source Detection: Iterative Approach

- Can be used when single-source model is not valid
  - Overlapping/correlated (in terms of data) sources
- Fit *n*-source model and determine the distribution of residual data
  - $$\Pr(\mathbf{R}_n|\mathbf{D}, H_n) = \int \Pr(\mathbf{R}_n|\Theta, H_n) \Pr(\Theta|\mathbf{D}, H_n)\, d\Theta$$
- Analyse residual data and compare between:
  - H_0 = "there is no additional source, residual data is due to noise only"
  - H_1 = "there is an additional source present"
- If H_1 is preferred then fit for $n+1$ sources and repeat the procedure
- Example: Extra-solar planet detection
  - See Feroz, Balan & Hobson, 2011, arXiv:1012.5129

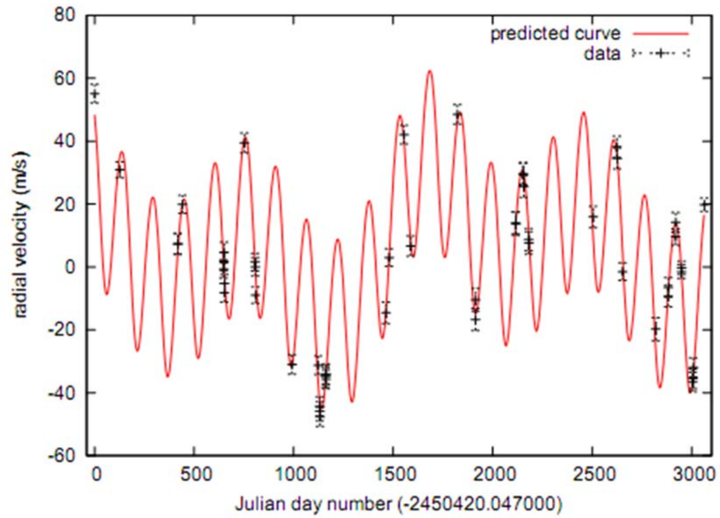
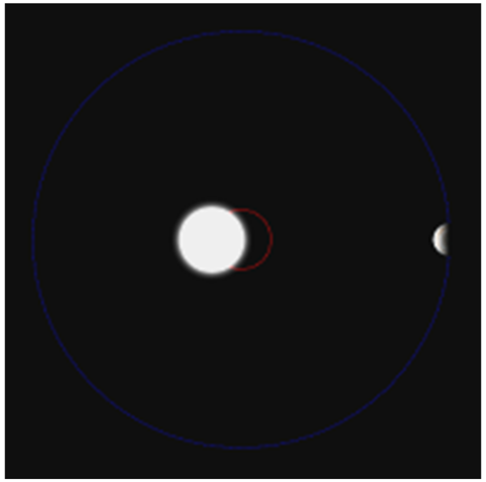


Figure 1. Radial velocity measurements, with 1σ errorbars, and the mean fitted radial velocity curve with three planets for HD 37124.

Applications: Exoplanet Detection

$$v(t_i, j) = V_j - \sum_{p=1}^{N_p} K_p \left[\sin(f_{i,p} + \omega_p) + e_p \sin(\omega_p) \right]$$



where

V_j = systematic velocity with reference to j^{th} observatory

K_p = velocity semi-amplitude of the p^{th} planet

ω_p = longitude of periastron of the p^{th} planet

$f_{i,p,}$ = true anomaly of the p^{th} planet

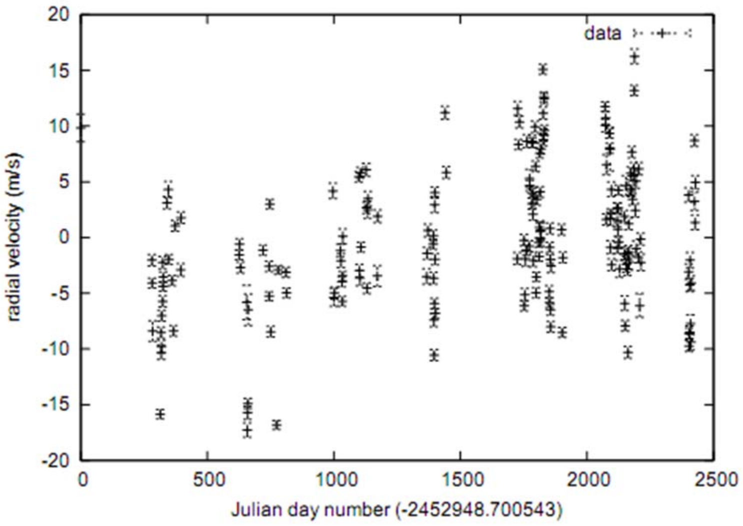
e_p = orbital eccentricity of the p^{th} planet

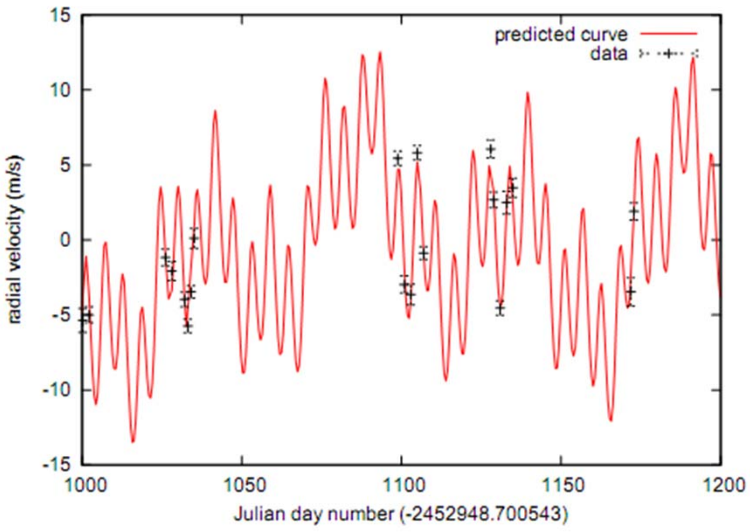
e_p = orbital period of the p^{th} planet

e_p = fraction of an orbit of the p^{th} planet, prior to the start of data taking at which periastron occurred

Applications: Exoplanet Detection – HD 10180

Feroz, Balan & Hobson, 2011, arXiv:1012.5129





Parameter	HD 10180 b	HD 10180 c	HD 10180 d	HD 10180 e	HD 10180 f	HD 10180 g
P (days)	5.76 ± 0.02	16.35 ± 0.05	49.74 ± 0.20	122.75 ± 0.54	600.17 ± 13.75	2266.22 ± 412.42
	(5.76)	(16.36)	(49.74)	(122.69)	(601.88)	(2231.44)
K (m/s)	4.54 ± 0.12	2.89 ± 0.13	4.28 ± 0.14	2.91 ± 0.14	1.43 ± 0.20	3.06 ± 0.16
	(4.63)	(2.94)	(4.25)	(2.70)	(1.79)	(2.98)
e	0.07 ± 0.03	0.13 ± 0.04	0.03 ± 0.02	0.09 ± 0.04	0.15 ± 0.09	0.09 ± 0.05
	(0.08)	(0.12)	(0.03)	(0.08)	(0.25)	(0.05)
ϖ (rad)	2.60 ± 0.38	2.62 ± 0.35	2.56 ± 0.16	2.65 ± 0.53	3.08 ± 0.97	2.89 ± 2.60
	(2.51)	(2.49)	(5.12)	(2.95)	(2.43)	(5.98)
χ	0.22 ± 0.06	0.35 ± 0.06	0.43 ± 0.27	0.23 ± 0.11	0.31 ± 0.28	0.67 ± 0.10
	(0.24)	(0.37)	(0.83)	(0.16)	(0.27)	(0.73)
$m \sin i$ (M_J)	0.04 ± 0.00	0.04 ± 0.00	0.08 ± 0.00	0.07 ± 0.00	0.06 ± 0.00	0.20 ± 0.01
	(0.04)	(0.04)	(0.08)	(0.07)	(0.07)	(0.20)
a (AU)	0.06 ± 0.00	0.13 ± 0.00	0.27 ± 0.00	0.49 ± 0.00	1.42 ± 0.03	3.45 ± 0.16
	(0.06)	(0.13)	(0.27)	(0.49)	(1.42)	(3.40)

Applications: Exoplanet Detection – GJ 667C

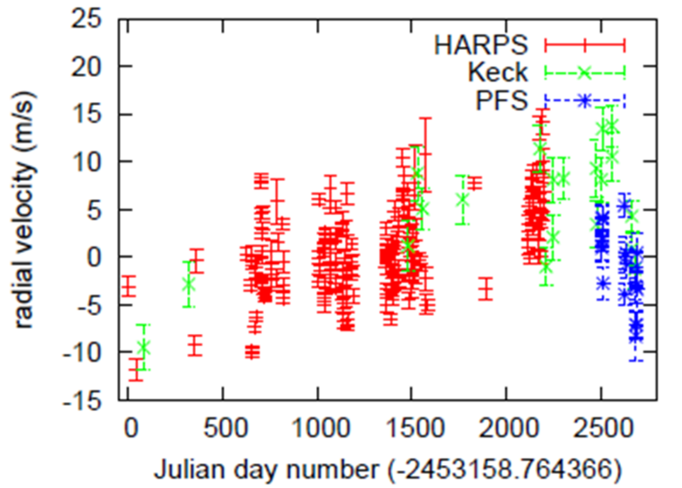
Feroz & Hobson, 2013, arXiv:1307.6984

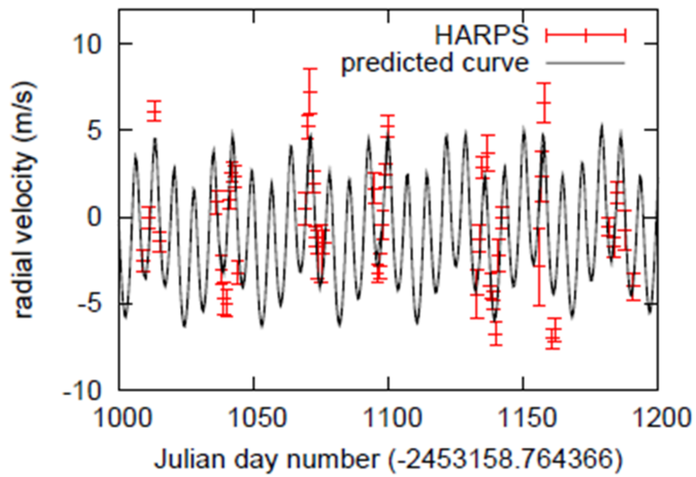
- Claims of up to 7 planets, with 3 super-Earths inside the habitable zone (arXiv:1306.6074)

- Bayesian analysis by Feroz & Hobson, with correlated 'red' noise and iterative approach found evidence for no more than 3 planets (arXiv:1307.6984)

N_p	$\mathcal{D}_{CCF}$ white noise	$\mathcal{D}_{CCF}$ red noise	$\mathcal{D}_{TERRA}$ white noise	$\mathcal{D}_{TERRA}$ red noise
1	17.05 ± 0.16	4.22 ± 0.16	16.95 ± 0.16	6.82 ± 0.16
2	9.80 ± 0.16	2.24 ± 0.15	18.94 ± 0.16	5.00 ± 0.16
3	2.57 ± 0.15	0.44 ± 0.14	4.22 ± 0.15	0.89 ± 0.15
4	0.13 ± 0.14	0.16 ± 0.14	1.37 ± 0.15	0.00 ± 0.15
5			-0.49 ± 0.14	

Log-evidence values for residual data (after detection of N_p planets) favouring 1-planet model over 0-planet model. $\mathbf{D}_{CCF}$ and $\mathbf{D}_{TERRA}$ are the radial velocity data-sets obtained using CCF and TERRA data reduction methods respectively.



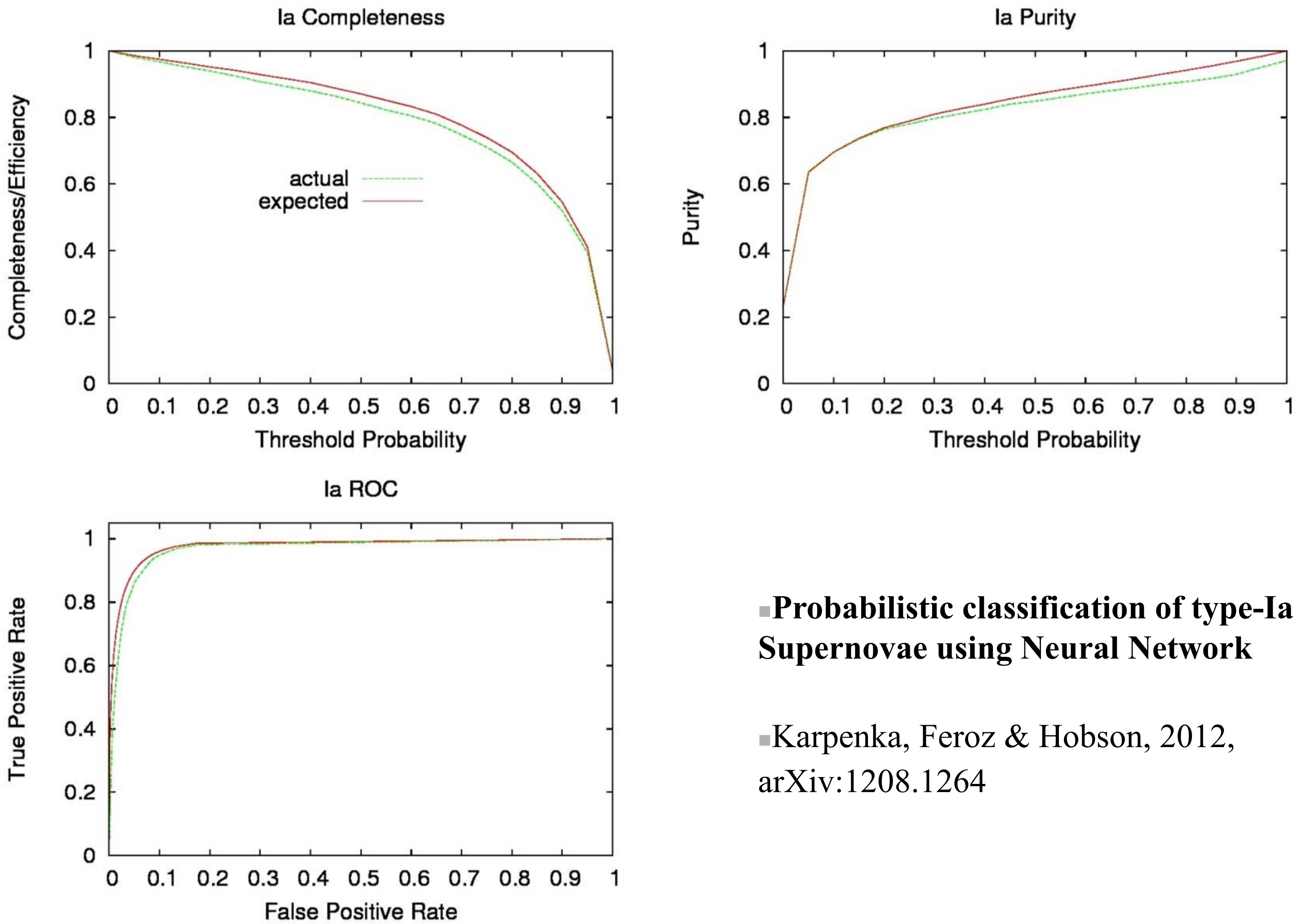


Top panel shows the observed radial velocity data. Bottom panel shows the mean fitted radial velocity curve overlaid on the observed data.

Purity, Completeness & Threshold Probability

- H_0 = "there is no source with its centre lying in the region S"
- H_1 = "there is one source with its centre lying in the region S"
- $$R = \frac{\Pr(H_1 \mid \mathrm{D})}{\Pr(H_0 \mid \mathrm{D})} = \frac{\Pr(\mathrm{D} \mid H_1)\Pr(H_1)}{\Pr(\mathrm{D} \mid H_0)\Pr(H_0)} = \frac{Z_1 \Pr(H_1)}{Z_0 \Pr(H_0)}$$
- $p_{TP} = R/(1+R)$
- Expected number of sources $\hat{N}_{tot} = \sum_{i=1}^{N} p_{TP,i}$
- Expected number of *true positives* $\hat{N}_{TP} = \sum_{i=1,\, p_{TP,i} > p_{th}}^{N} p_{TP,i}$
- Expected number of *false positives* $\hat{N}_{FP} = \sum_{i=1,\, p_{TP,i} > p_{th}}^{N} (1 - p_{TP,i})$
- Expected *completeness* $\hat{\varepsilon} = \hat{N}_{TP} / \hat{N}_{tot}$
- Expected *purity* $\hat{\tau} = \hat{N}_{TP} / (\hat{N}_{TP} + \hat{N}_{FP})$

Purity, Completeness & Threshold Probability



- **Probabilistic classification of type-Ia Supernovae using Neural Network**

- Karpenka, Feroz & Hobson, 2012, arXiv:1208.1264

Conclusions

- Bayesian statistics provide rigorous approach to astrophysical source detection
  - Use Bayesian model selection to distinguish real sources from spurious ones
- Efficient and robust source detection can be done using nested sampling
  - MultiNest allows sampling from multimodal/degenerate posteriors
  - local and global evidences and parameter constraints
  - typically ~ 100 times more efficient than standard MCMC
- Probabilistic source detection removes arbitrariness in choice of detection criterion
  - allows calculation of expected purity and completeness
- MultiNest publicly available
  - with SuperBayeS for SUSY phenomenology (www.superbayes.org)
  - as a standalone inference engine (www.mrao.cam.ac.uk/software/multinest)