Resolution, Regularisation and Revolutions

Brendon J. Brewer

School of Physics, The University of Sydney

Supervisor: A/Prof Geraint F. Lewis

Some Questions Involving Uncertainty

- Global warming – will skiing be possible in Australia in a few decades?
- Will England win the Ashes?
- Virtually any question in science that people are still debating – uncertainty exists

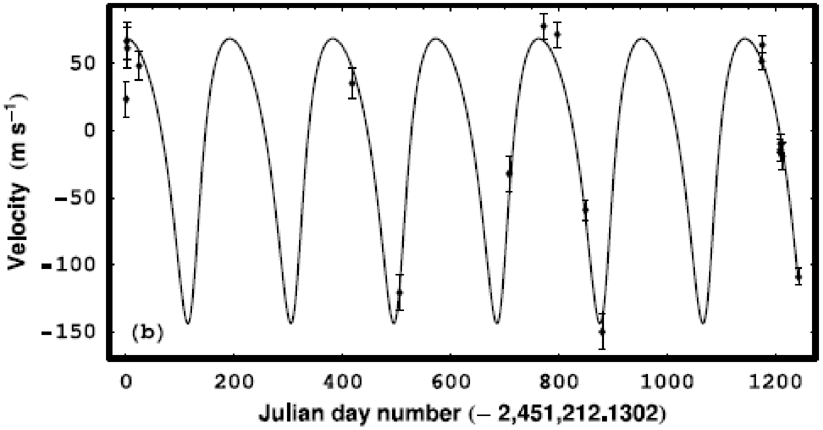
Astronomical Examples

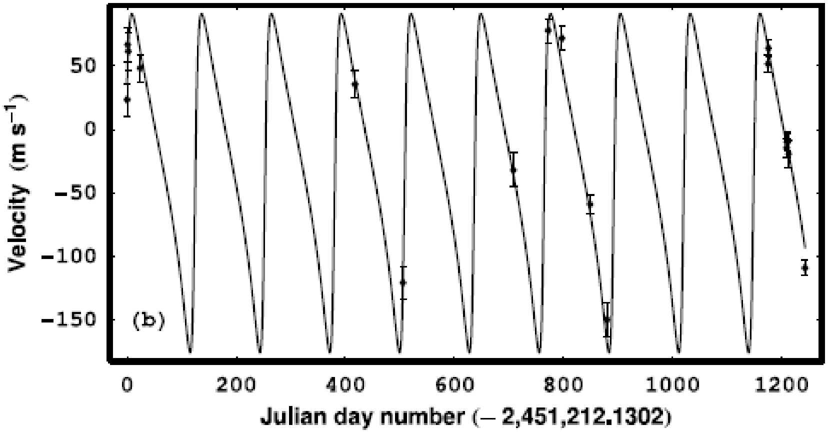
- Estimating cosmological parameters (e.g. from WMAP)
- Extrasolar planets (Gregory, 2005) – Estimate orbital parameters (mass,eccentricity etc)
- Any area of study involving uncertainty (e.g. your area!)

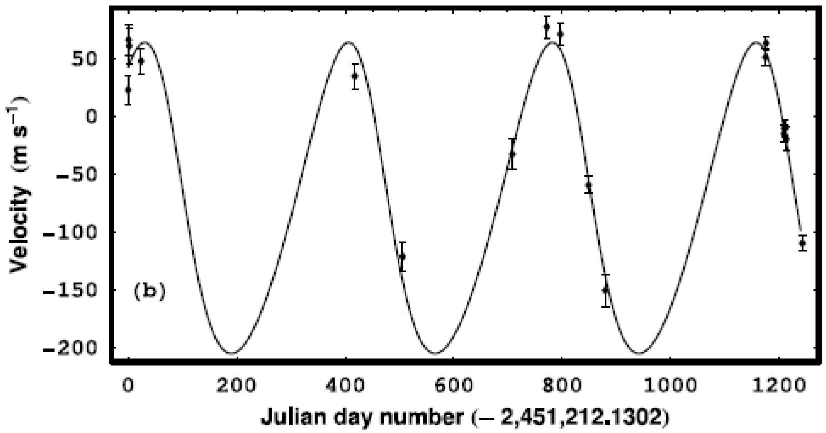
Motivation – How to Deal With Uncertainty?

- Cost (time, effort, money) of obtaining useful data is high. Make the most of your data!
- Any assumptions are explicit and easily criticised.
- All problems have the same method of solution. Physicists love unification.

Three Fits to the Same Data







Classical Statistics

- Distinguishes between "hypothesis testing" and "estimation" – use p-values and estimators respectively
- But isn't estimation really the testing of many hypotheses? $\theta=1, \theta=2, \theta=3, \theta=3.434$ etc.
- Isn't hypothesis testing really estimation? Have a set of hypotheses $\{H_i\}$, estimate i.
- Data are random, hypotheses are fixed

Bayesian Inference

- All probabilities are conditional probabilities
- Probabilities represent degrees of confidence. Millionaire
- The probability of things changes when the things you know change. P(rain|clouds), P(rain|no clouds)

P(clouds | France) > P(clouds | Australia)

(Insufficient data on Switzerland)

Estimation/Fitting

- It is conceptually simple to do Bayesian estimation.
- Assign prior distribution $p(\theta)$
- Data D updates this to the posterior distribution, proportional to $p(\theta)L(\theta;D)$

Simple Example

A ball in a bag is either black or white. A black ball is thrown in with it. A person draws a ball out at random and it is black.

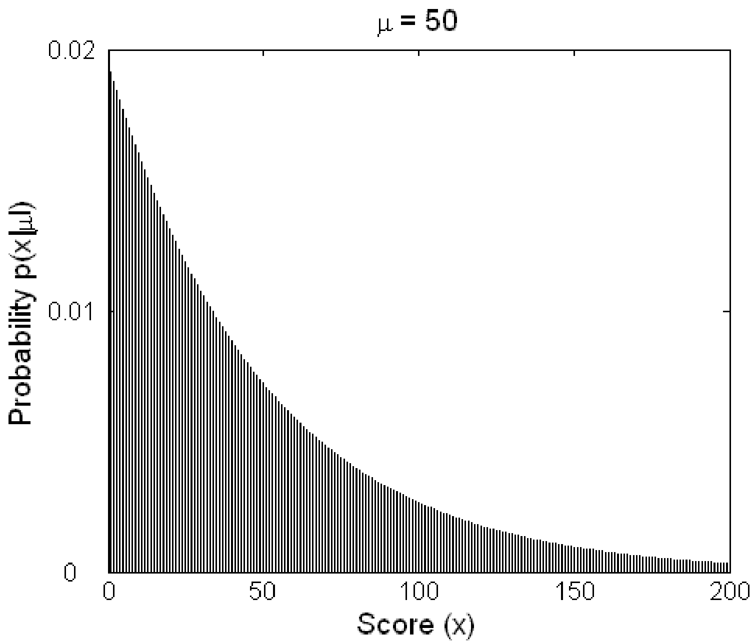
- Hypothesis 1: Ball was black
- Hypothesis 2: Ball was white

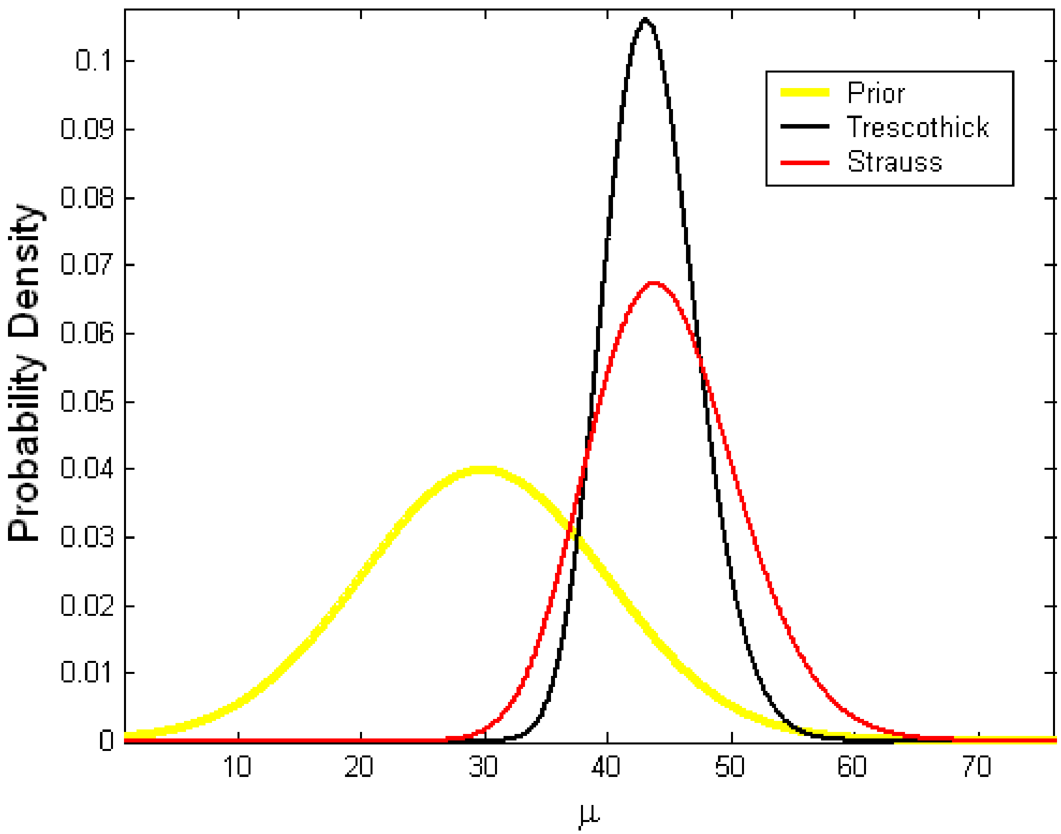
- $P(D|H_1) = 1$
- $P(D|H_2) = 1/2$

$$\frac{P(H_2|D)}{P(H_1|D)} = \frac{P(H_2)}{P(H_1)} \times \frac{P(D|H_2)}{P(D|H_1)}$$

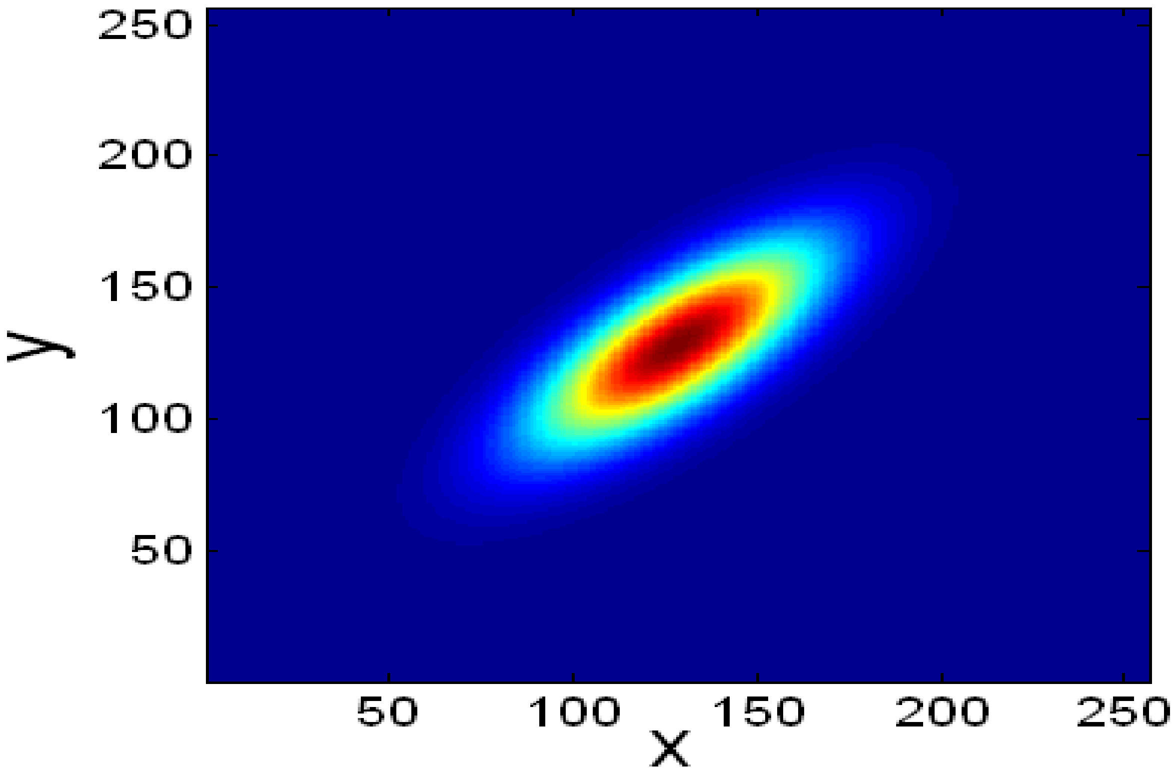
| not put in by these guys

Yes, you really do just multiply two functions





Marginalisation - delete



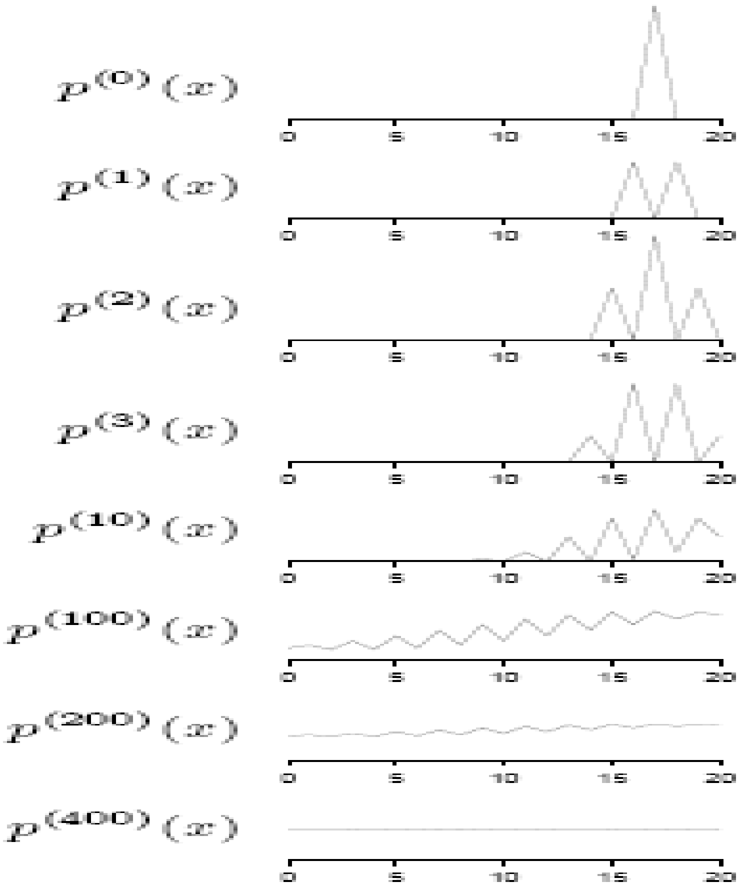
Chain Monte Carlo – Move to later (after defining problems) – maybe delete

- Usually, we can calculate a value proportional to the density at x for any point x

- Metropolis algorithm: Propose small change $x \rightarrow x'$, accept with probability

$$\text{Min}(1, p(x')/p(x))$$

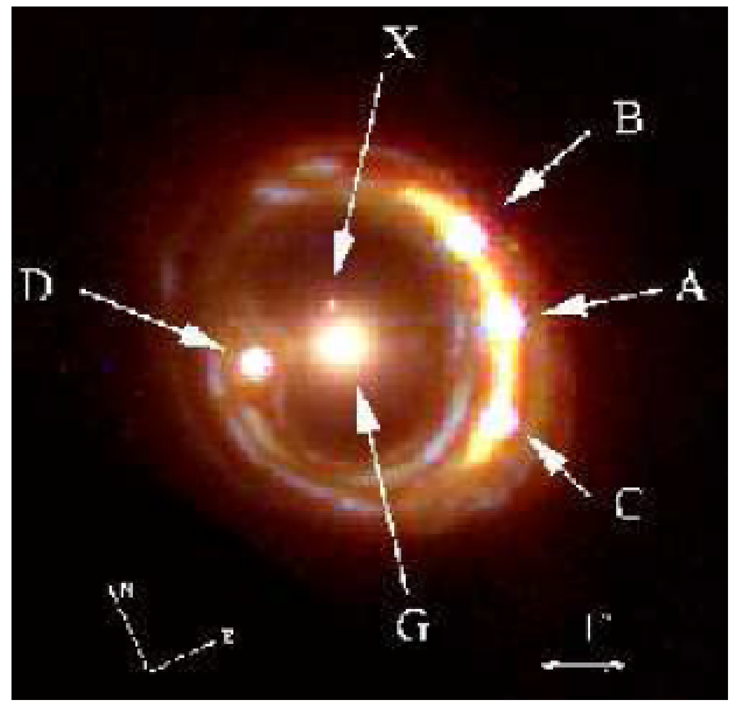
MCMC converging to a target distribution



From “Information Theory, Inference and Learning Algorithms”, by David Mackay

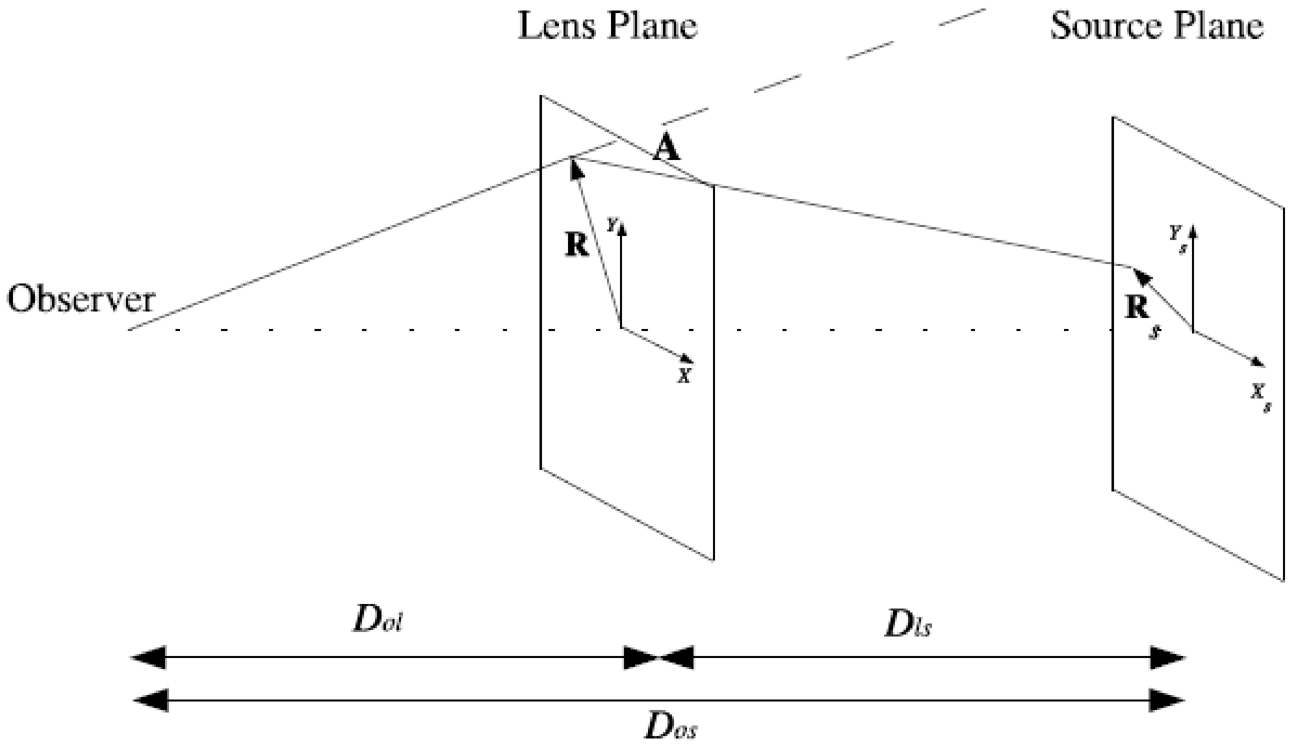
Gravitational Lens Inversion

- Use gravitational lens as a “natural telescope”



ER 0047-2808 (source at redshift 3.6)   J1131

Lensing Basics



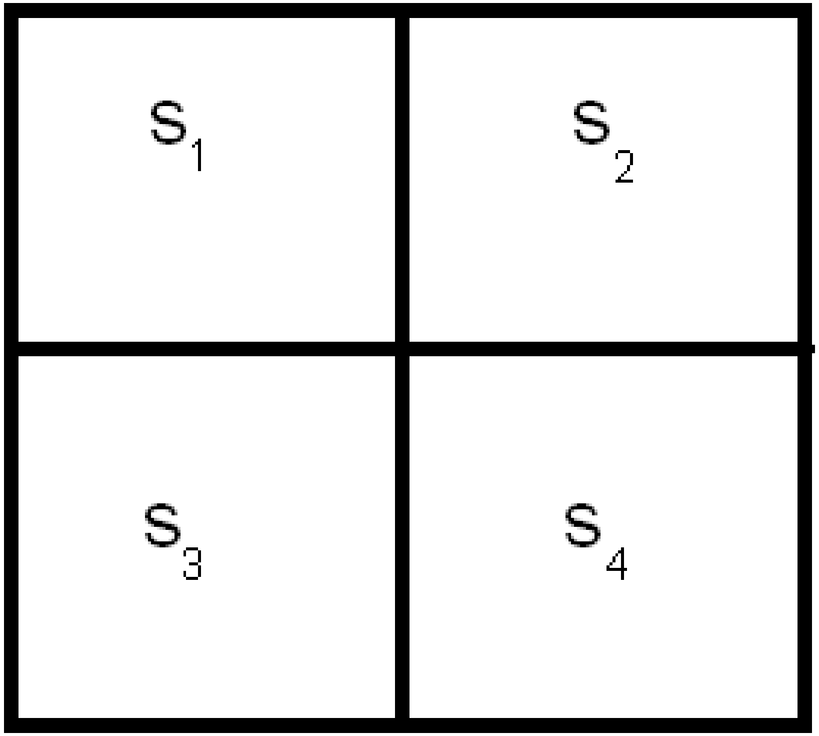
The Likelihood Function for Gaussian Noise

- Independent error for each pixel of the image

- Joint density for the image given the lens and source is the product of N Gaussians, one for each pixel

- It is proportional to $\exp(-0.5*\chi^2)$

Pixellated Sources



Least Squares

- Given the observed image, want to estimate the true source profile (and lens mass distribution parameters)

- Find those source pixel values and lens parameters that make χ^2 a minimum

- Relatively fast (few hours!!!)

Problems with Least Squares

- Usually leads to negative pixels
- A non-unique solution is possible, especially if we try to use a lot of pixels
- Get spiky solutions due to PSF
- ***"Single answer" syndrome***

Bayesian Interpretation of Least Squares

- Log(posterior) = log(prior) + log(likelihood)
- Likelihood proportional to exp(-0.5*χ^2)
- χ^2 minimisation is equivalent to finding the posterior mode for a uniform prior. Do we really want a uniform prior over pixel values?

Constrained Least Squares

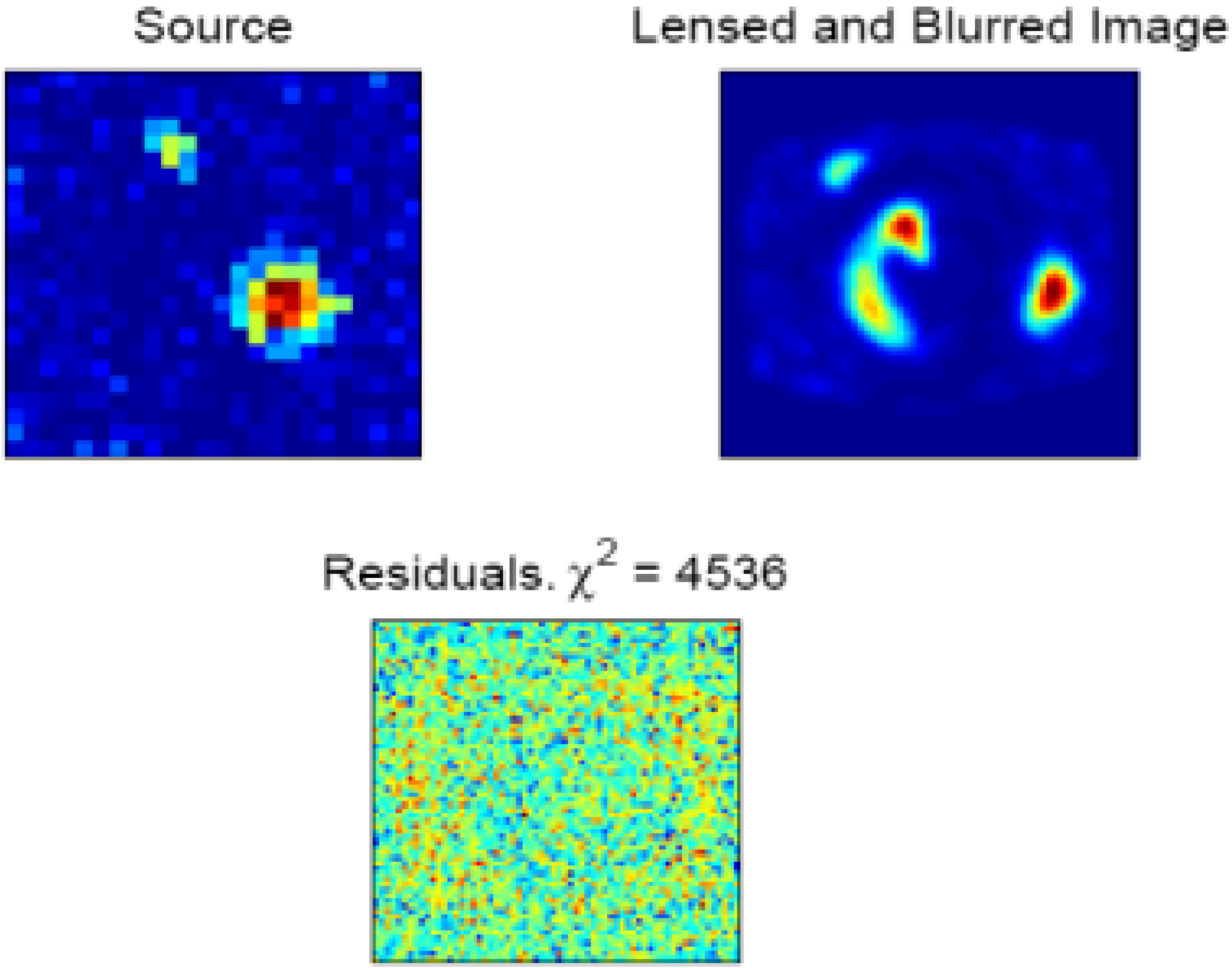
- Maximise likelihood under the constraint that each pixel value is nonnegative
- We used genetic algorithms to do this
- ***"Single answer" syndrome***

Problems With Constrained Least Squares

- Can be interpreted as Bayesian prior that is uniform over the space of possible sources

- But this prior isn't really a good description of our prior knowledge

- Leads to incorrect estimates!! Bright sky


Source
Lensed and Blurred Image
Residuals. χ^2 = 4536

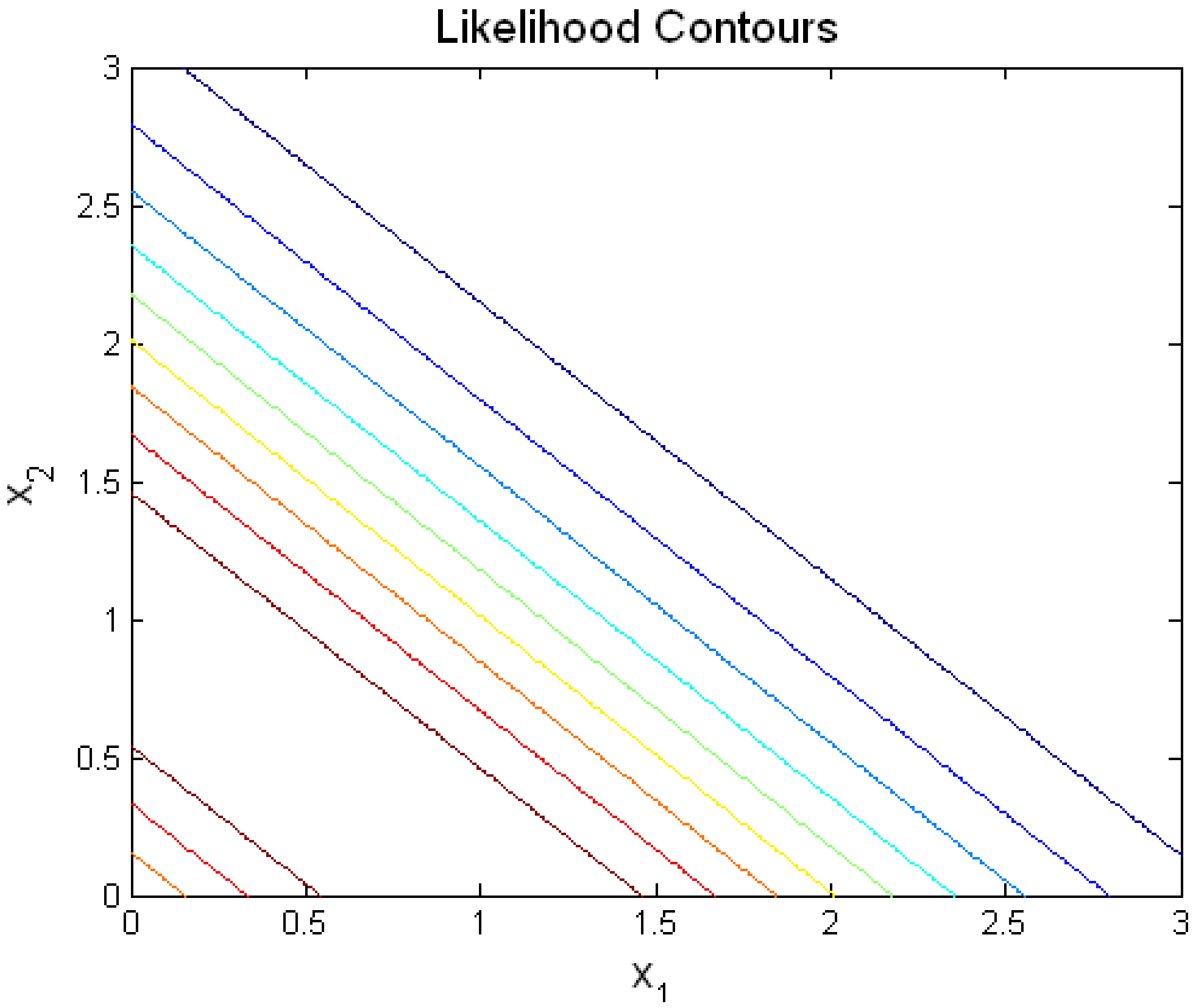
A Toy Example

- Take a noisy measurement of a number that is supposed to be the sum of two positive numbers x_1 and x_2.

- Get data D. $D|(x_1,x_2) \sim$ Gaussian with a mean $x_1 + x_2$ and standard deviation 1.

- Observe D=1. What do we know about x_1 and x_2? In particular, what is the probability that $x_1 + x_2 > D$?

- $p(D|x1,x2) = C*\exp[-0.5*(D - (x^1+x^2))^2]$



Regularisation

- Rather than maximising $-0.5\chi^2$, maximise

$$Q(s) - 0.5*\chi^2$$

Q(s) is called the regulariser. Usually chosen to penalise “spikiness”

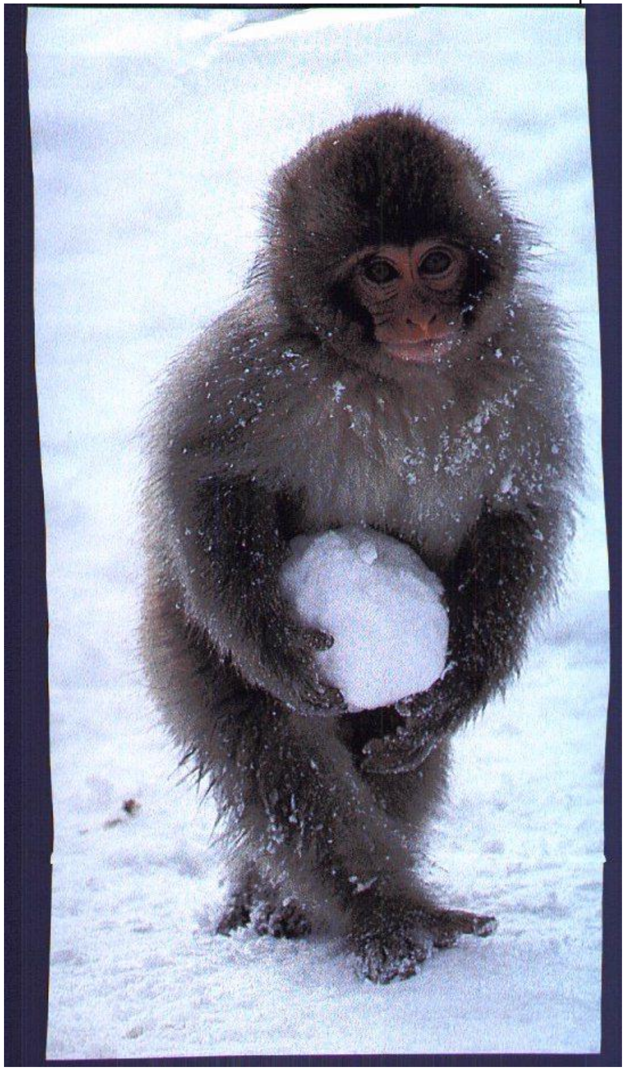
- ***“Single answer” syndrome***

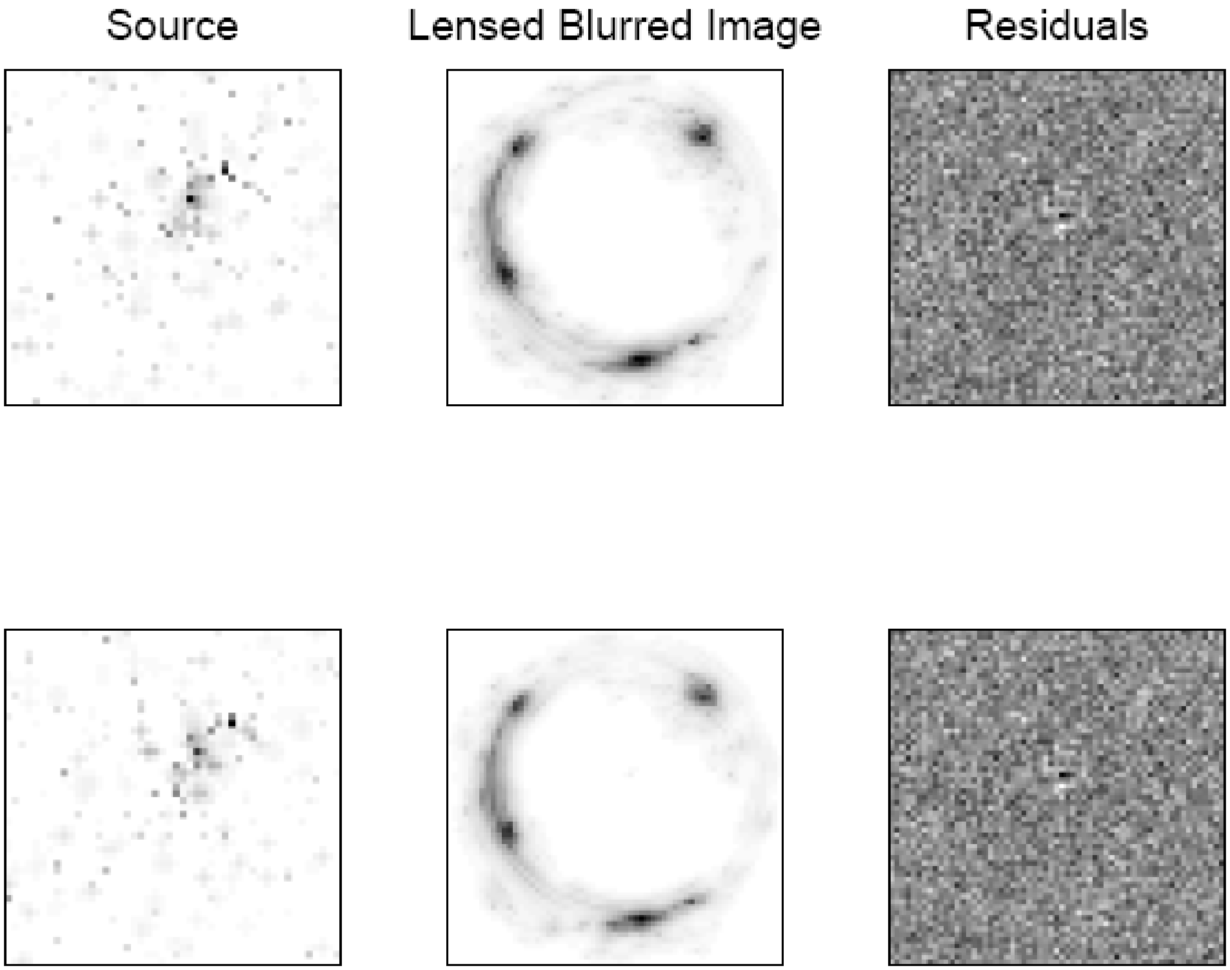
Why You Need a Prior

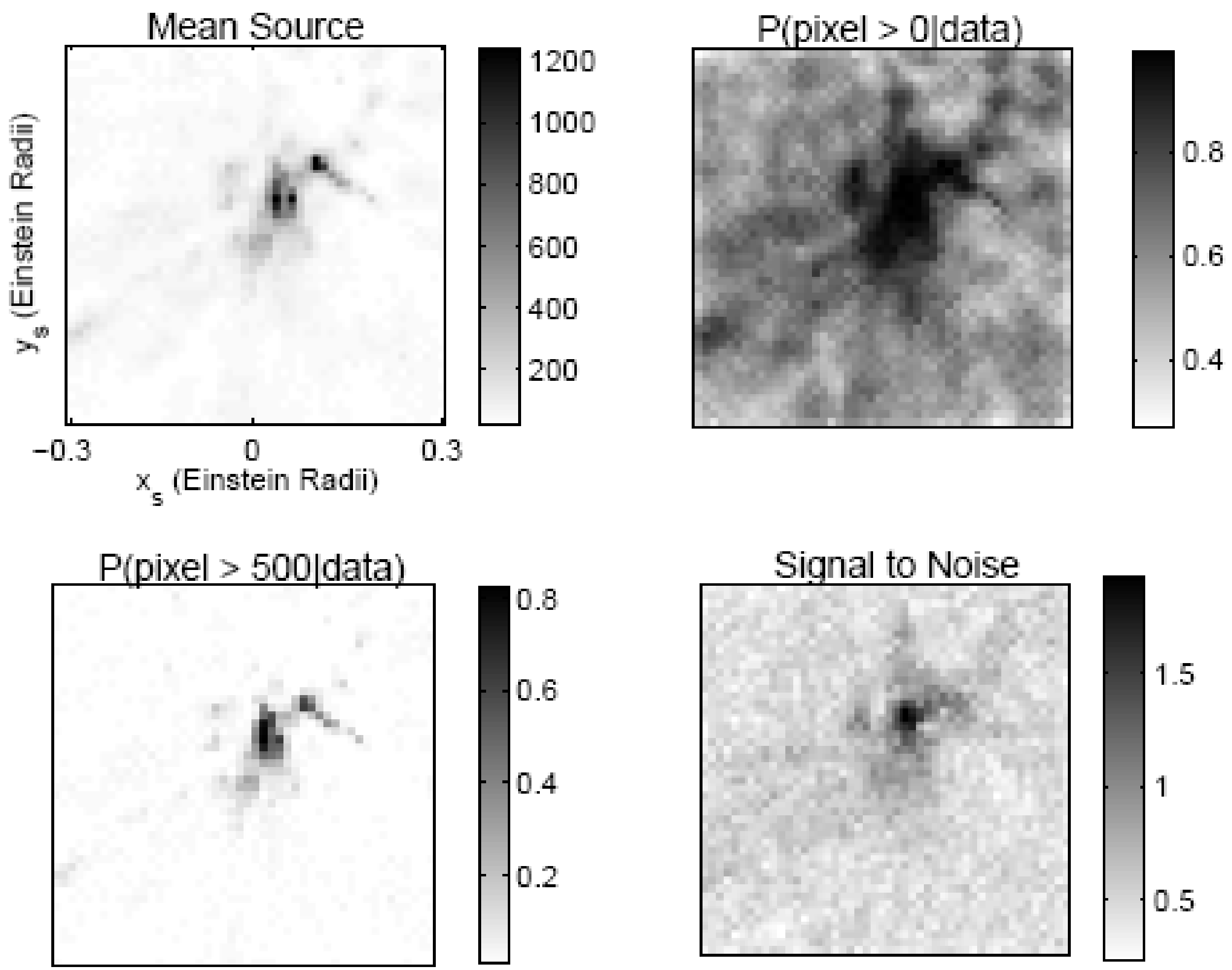
- A whole set of models is consistent with the data
- The diversity of this set of models → uncertainty
- Plenty of ridiculous models fit the data well → need to have some measure of prior plausibility.

Our Prior

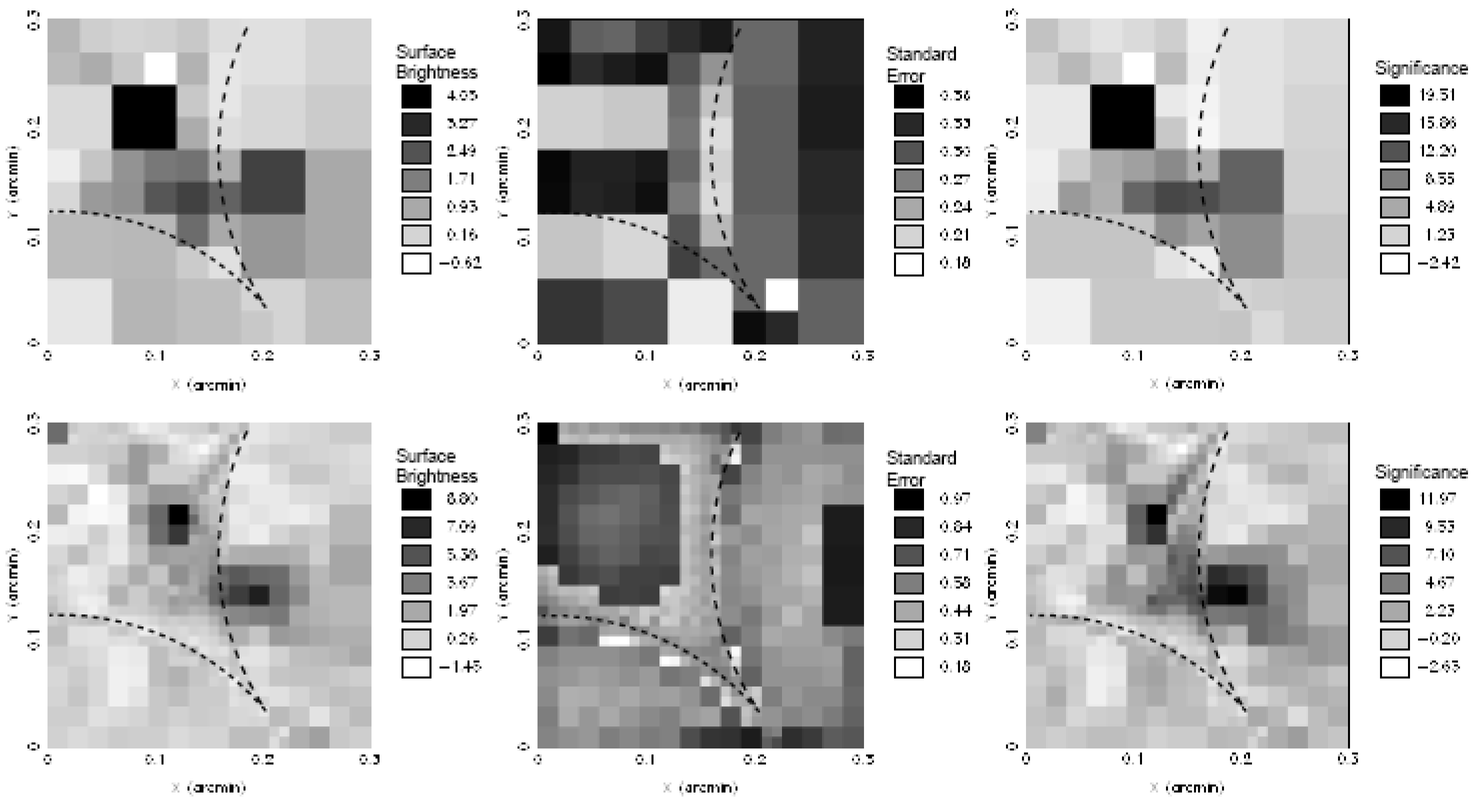
Multiscale Monkey Prior




Source
Lensed Blurred Image
Residuals


Mean Source
1200
1000
800
600
400
200
y_s (Einstein Radii)
−0.3
0
0.3
x_s (Einstein Radii)
P(pixel > 0|data)
0.8
0.6
0.4
P(pixel > 500|data)
0.8
0.6
0.4
0.2
Signal to Noise
1.5
1
0.5

Regularised Result (Dye and Warren)



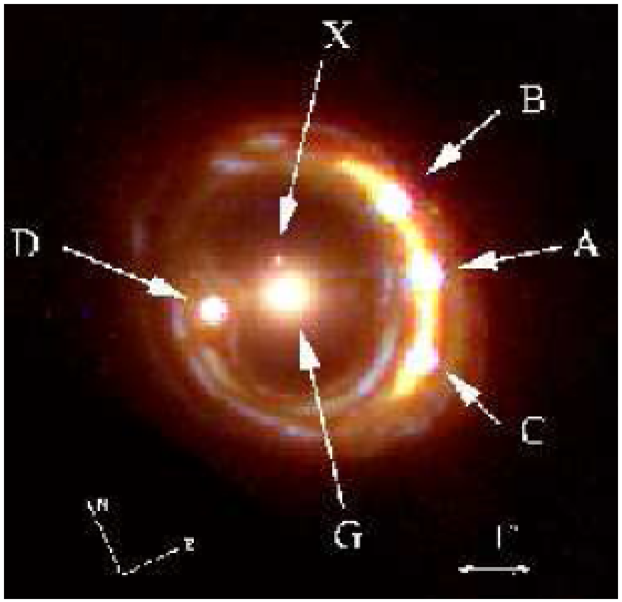
Tricks to Increase Efficiency - Delete

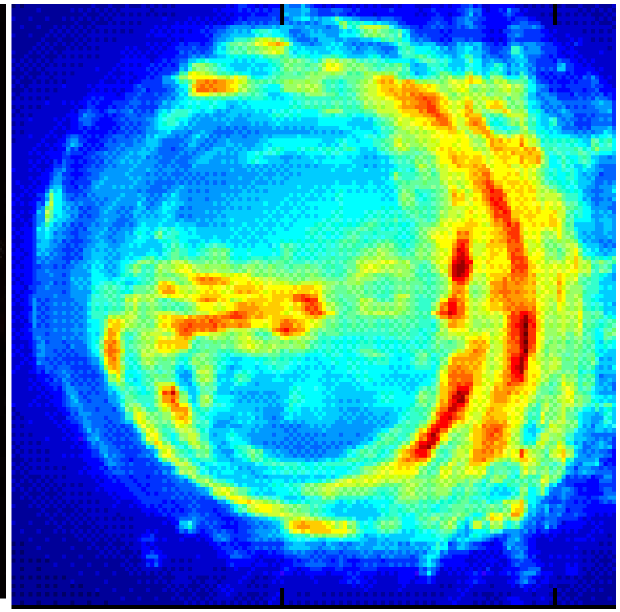
- Many modifications to standard MCMC exist, designed to handle multimodality, correlations, etc

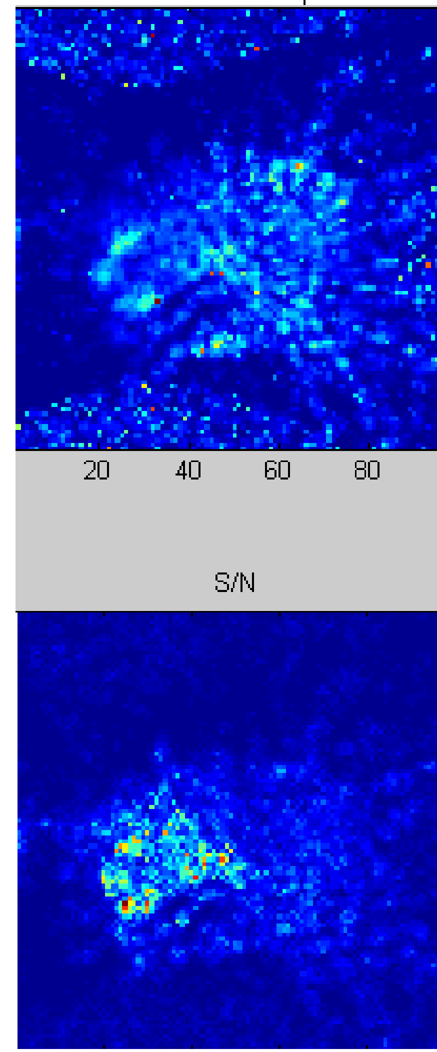
- Tempered Transitions – raise likelihood to a power < 1, makes exploration easier. Decrease this power then increase it

Why is our result better

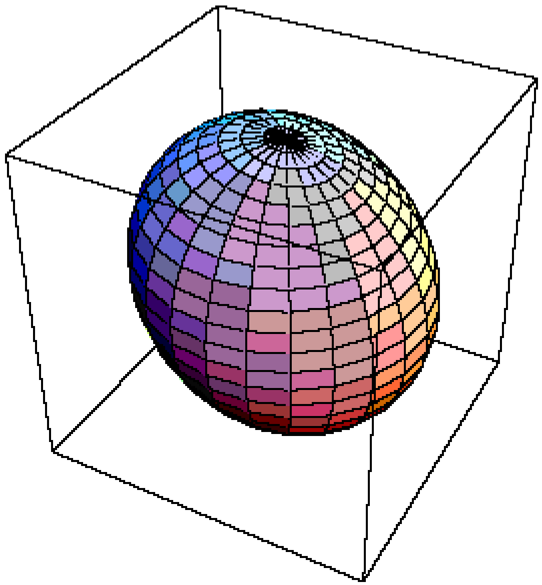
- Achieved higher resolution
- This was only possible because the prior was actually chosen as a model of prior knowledge
- With many pixels (needed for high resolution), the choice of regulariser becomes important, so it better be a good prior


X
B
D
A
C
G




20
40
60
80
1
S/N

Asteroseismology

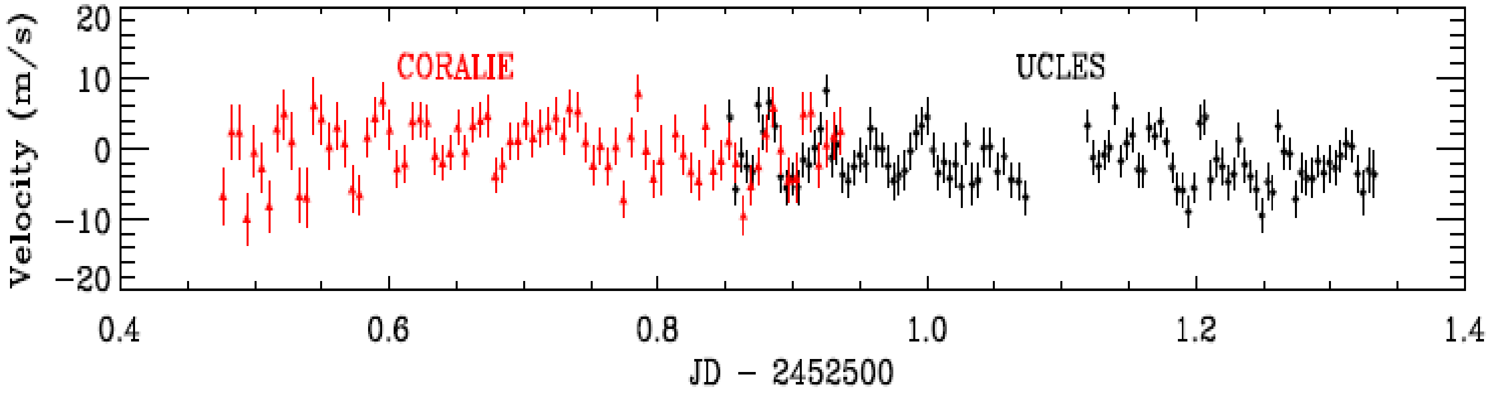


Like listening to the sound of a musical instrument

...and inferring the physical structure from the sound.

Asteroseismology

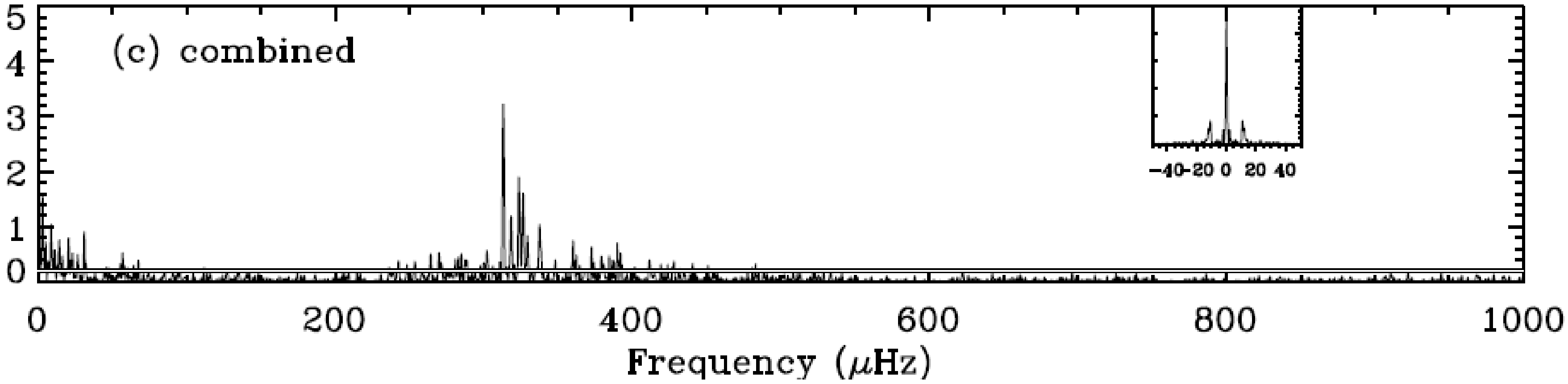
- Signal consists of sine waves plus noise
- Want to know the frequencies. The pattern of frequencies depends on stellar parameters



Power Spectrum/Fourier Transform

- Dot product of data with a sine/cosine wave, as a function of frequency. Peaks at the value of the frequency

- Window function – Like PSF



CLEANing

- Fit sine wave by least squares, then subtract
- Keep going until "only noise" remains
- Gives an estimate of the frequency pattern.
- ***"Single answer" syndrome***

Bayesian Approach

- Fit of a nonlinear model

$$y(t) = \sum_{j=1}^{M} \left(A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t) \right)$$

- Estimate unknown parameters m, {A}, {B}, {f}. The amplitudes are nuisance parameters
- Can integrate them out **analytically**

What happens?

- Bretthorst (1988) – Power spectrum is proportional to the log of the posterior density for a frequency (after some approximations)

Our Approach

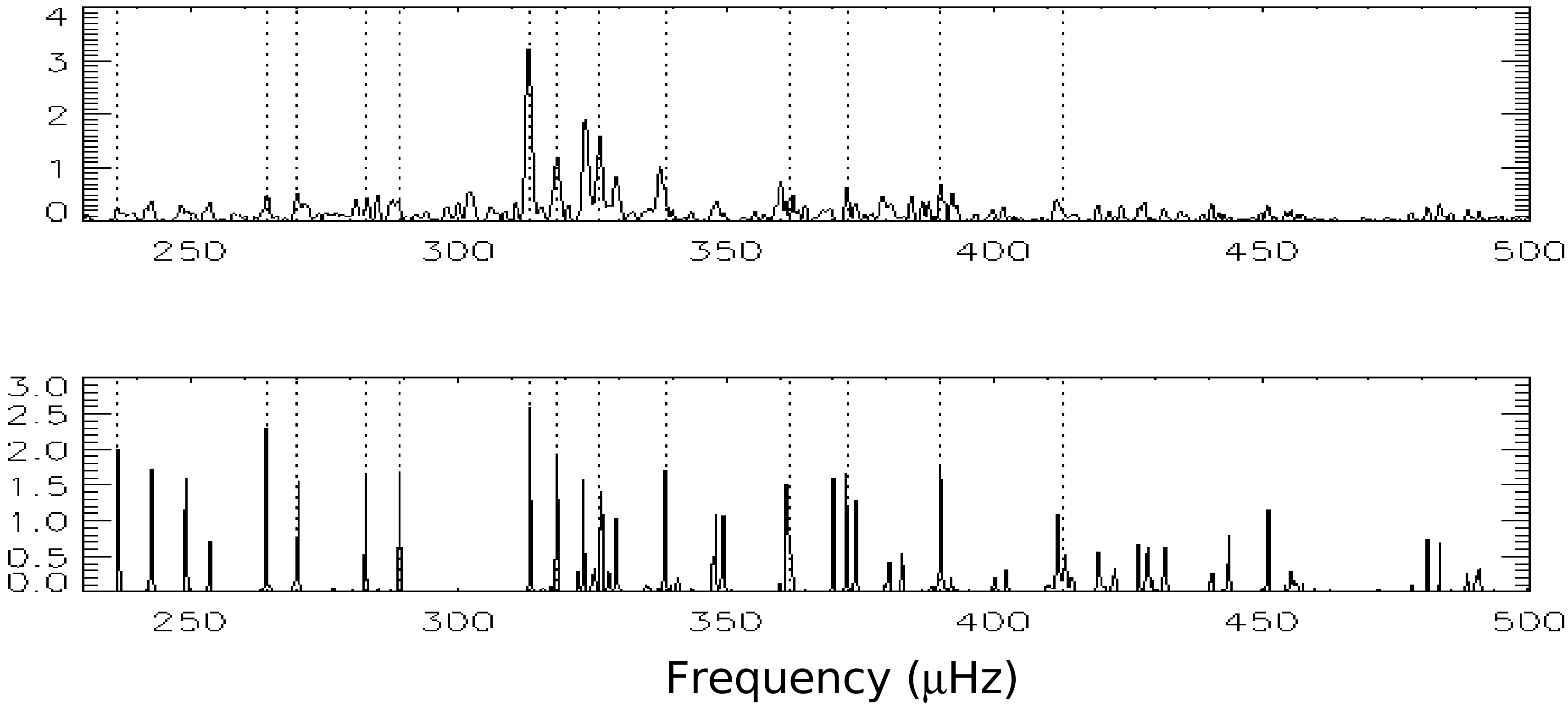
- Get a whole sequence of plausible models

- Using Markov Chain Monte Carlo algorithm (useful computational method for exploring high dimensional parameter spaces)

- Diversity of models gives uncertainties

Application to ν Indi

- Power spectrum on top
- Bayesian frequencies below (area proportional to detection confidence)
- Dotted lines – CLEAN frequencies



Conclusions

- A consistent, simple approach to statistical analysis exists
- Gives sensible results to the inverse problems tested here
- Improvements were obtained over standard methods
- Sometimes will get major improvements, sometimes get the same result, but it's worth trying