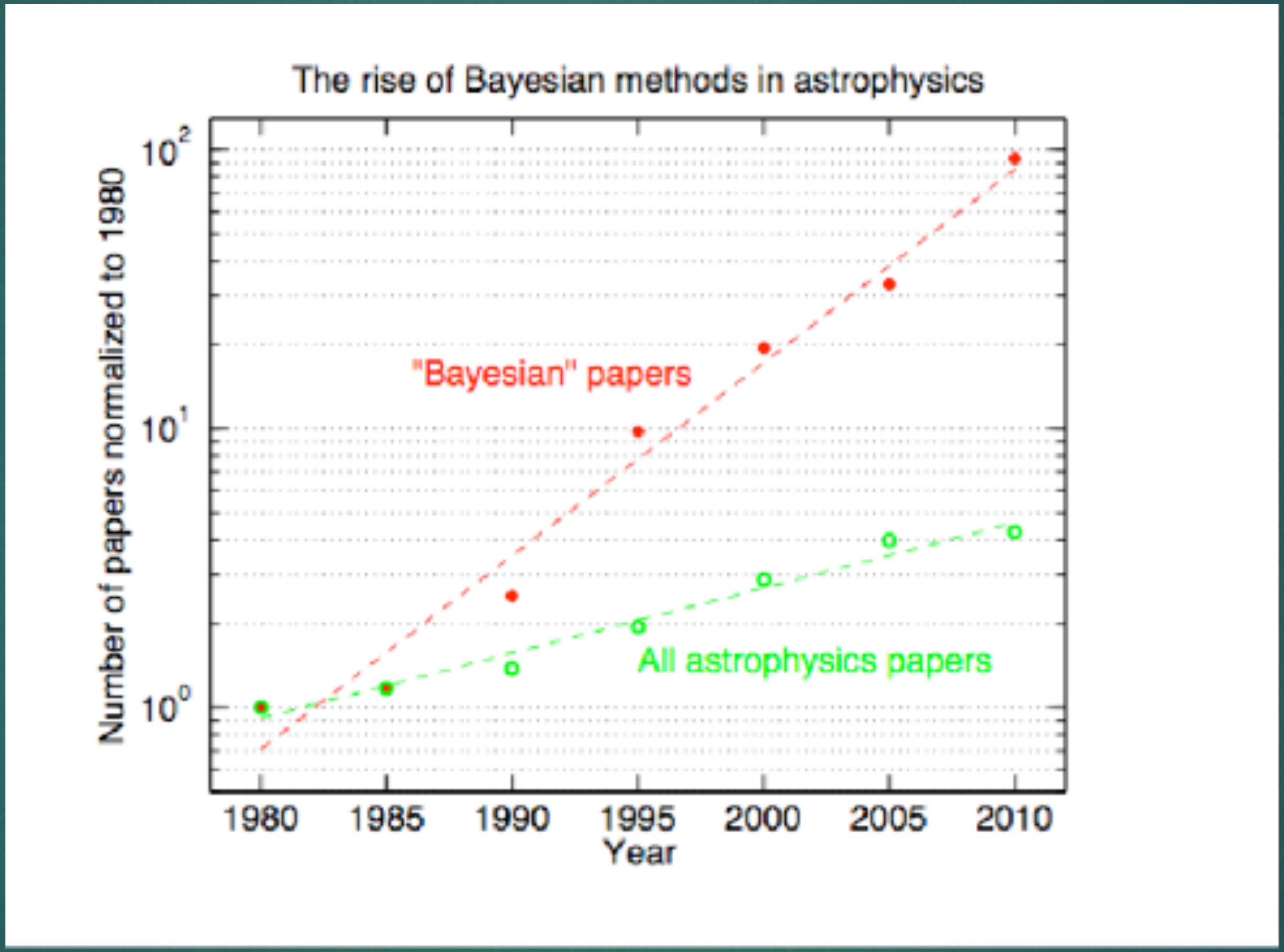
Bayesian Techniques In Pulsar Timing



LINDLEY LENTATI
CAMBRIDGE UNIVERSITY

Overview

Part 1:

Baye’s Theory - Terminology
Methods
Examples

Part 2:

Noise in pulsar timing - Intrinsic (Glitches, Timing Noise)
The Interstellar Medium
Systematics
The Solar System
…

Part 1:
Bayesian Statistics

Bayesianism and Frequentism

Start at the heart of it..

Asks two different questions:

Frequentist:

What is the probability of my data, given my model?

Assumes model is fixed – data random variable

Bayes:

What is the probability of my model, given my data?

Assumes data is fixed – model is random variable

Bayes Equation

The Prior

The Likelihood

$$P(\theta \mid D) = \frac{P(\theta)P(D \mid \theta)}{\int P(\theta)P(D \mid \theta)d\theta}$$

The Posterior

The Evidence

Likelihood

P(D | M) : Probability of the data given the model

(The frequentist bit)

Most typically just Gaussian chi-sq:

E.g. for independent data points:

d = data

m = model

o = error on the data

P(D | M) = exp(-0.5(d-m)^2/o^2)

Prior

P(M) : The probability of our model parameters before we do the experiment.

Many different choices:

Uniform in the parameter

Uniform in the log of the parameter

Gaussian with mean and error

+ …

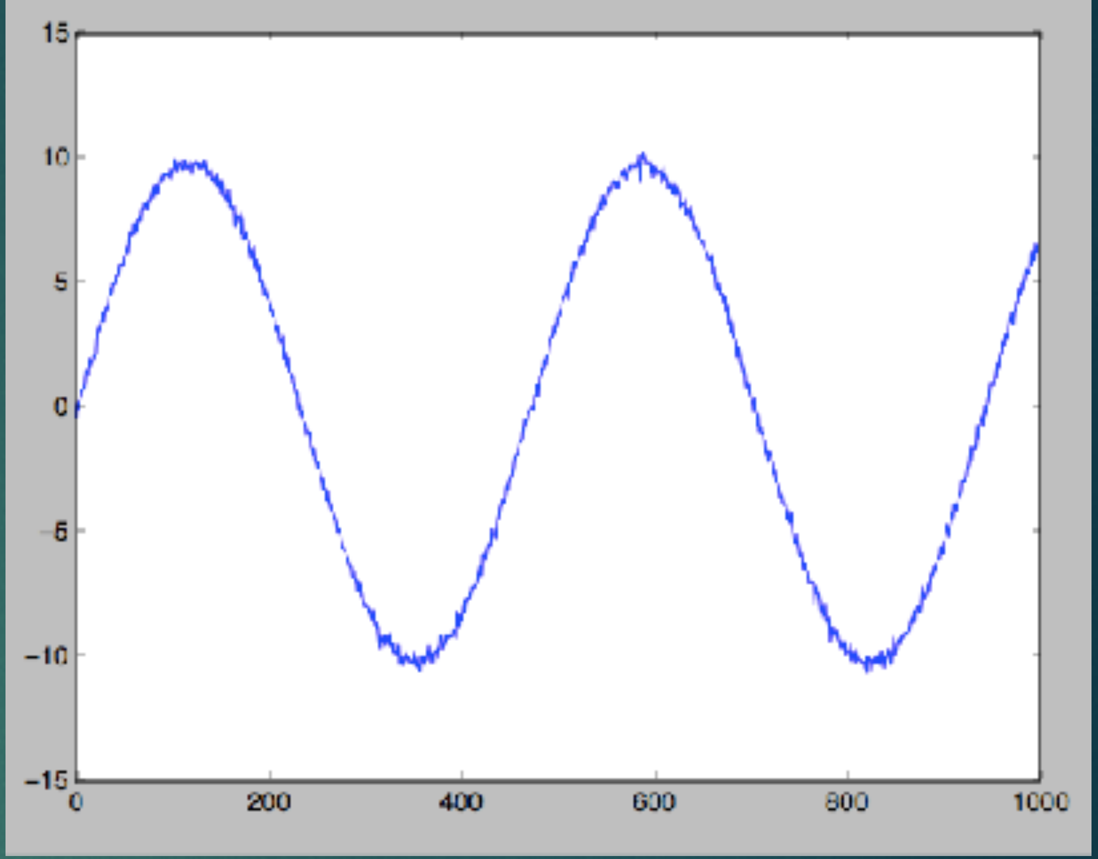
Prior

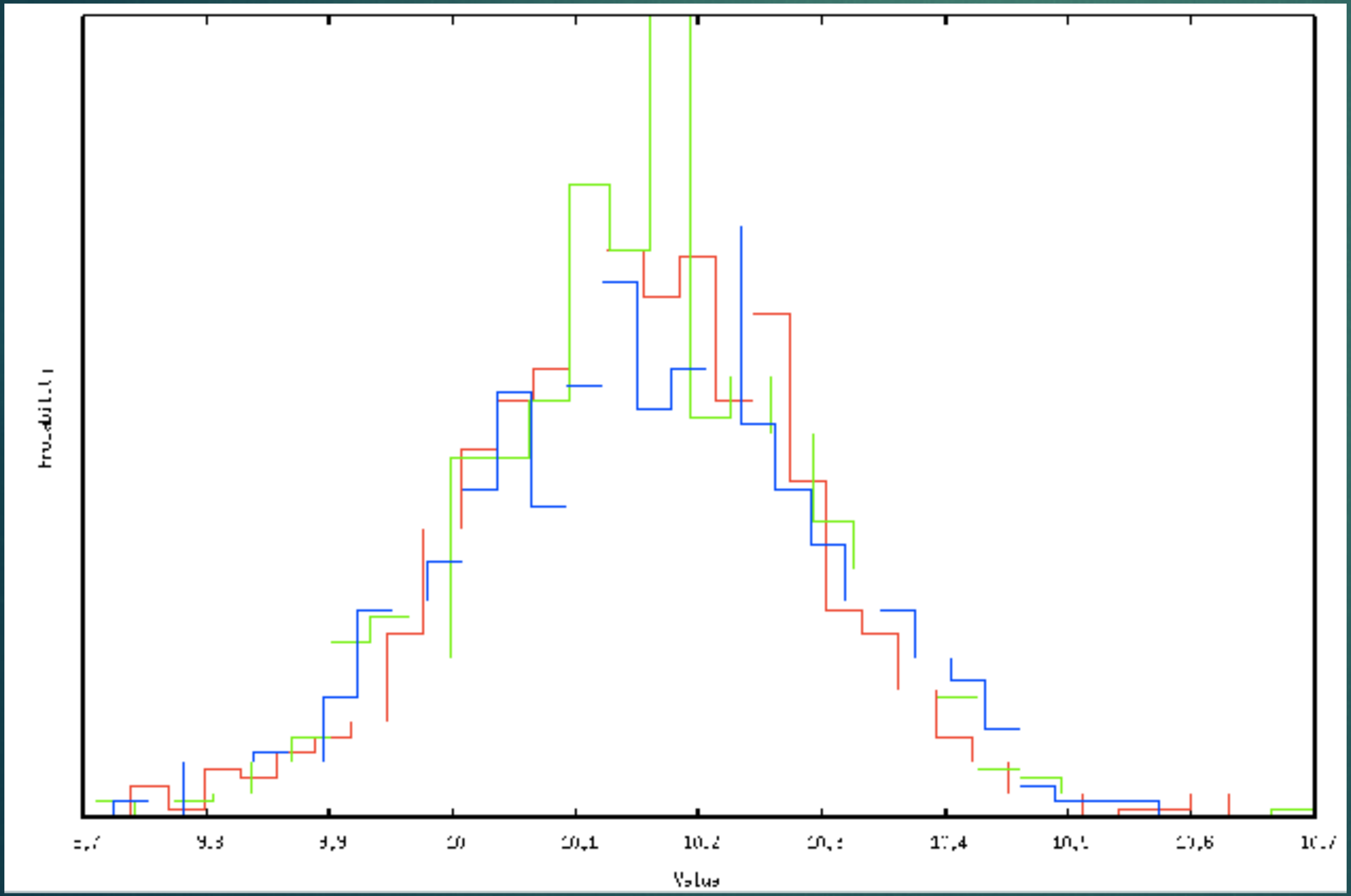
Different priors can lead to very different results.

Consider fitting for 1-dim problem: amplitude of sine wave (10) in some noisy data.

Not usually a problem in high signal-to-noise cases.

Here the data can update our current knowledge:





3 Priors:
Uniform in amplitude (red)
Uniform in log-amplitude (green)
Gaussian (blue)

Results are consistent:
E.g. consider Gaussian prior:
9.5 +/- 5

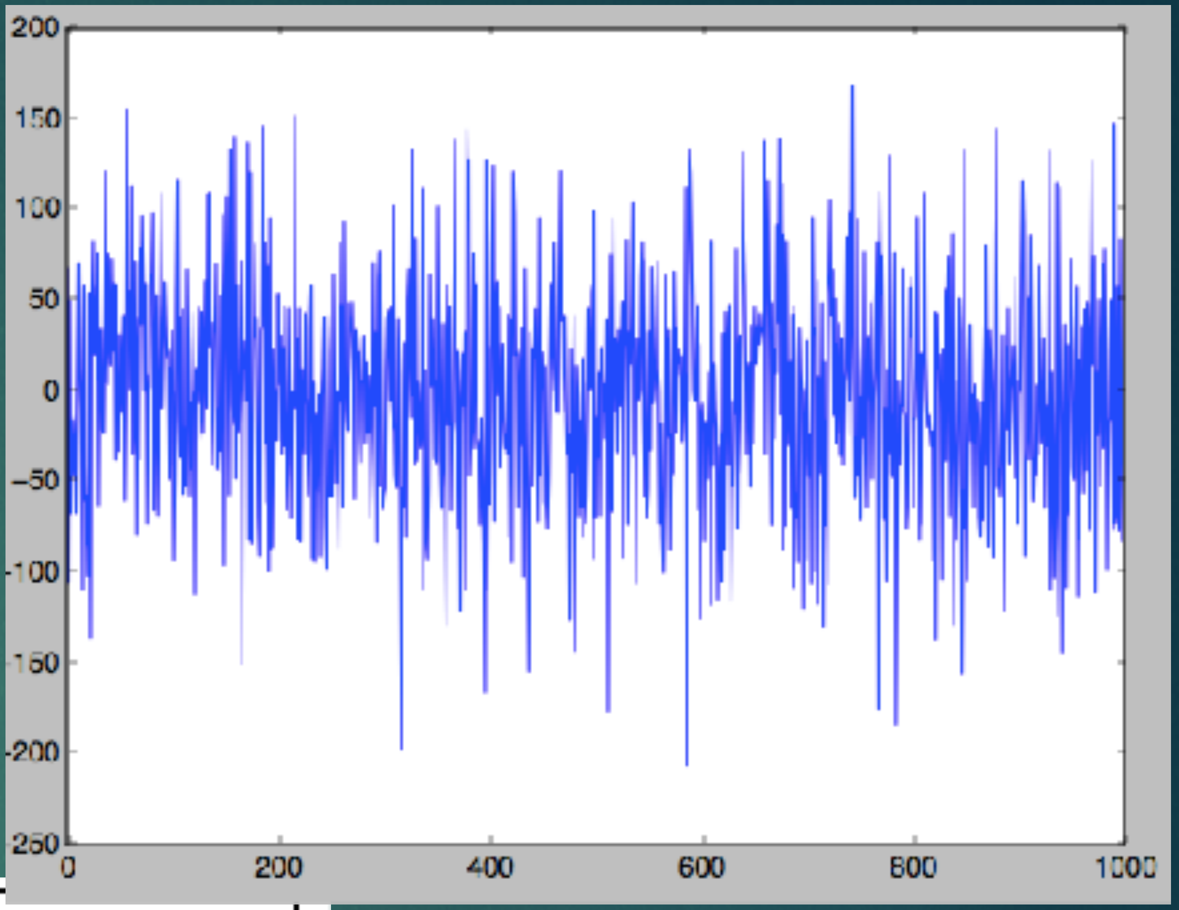
Probable range in posterior is
10.17 +/- 0.14

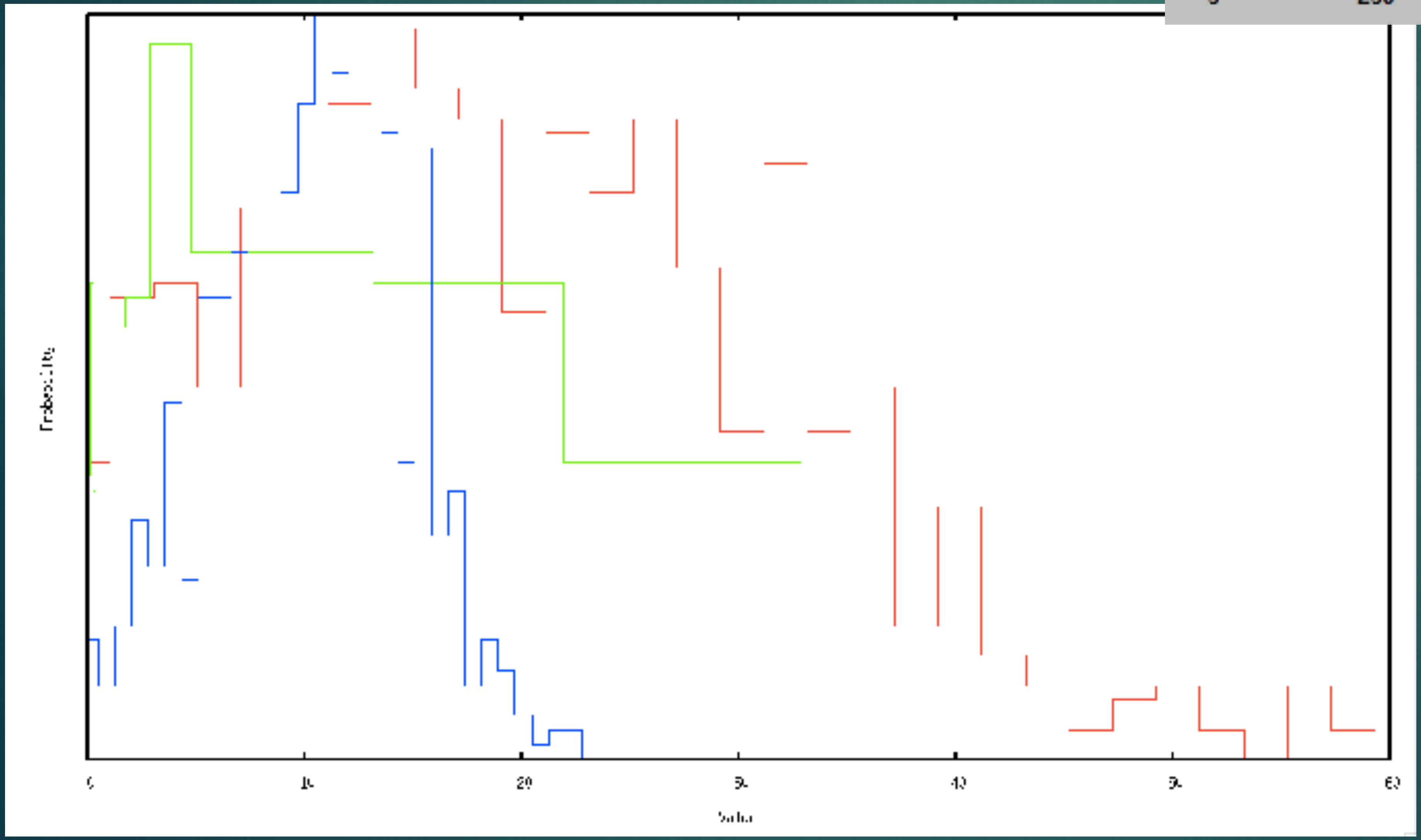
Prior decreases log-likelihood by 1 for change in parameter value that is huge compared to that inferred by data

Prior

Different priors can lead to very different results.

In the Low signal-to-noise case things are not so simple.





As before
3 Priors:
Uniform in amplitude (red)
Uniform in log-amplitude (green)
Gaussian (blue)

Now results not consistent

The Evidence

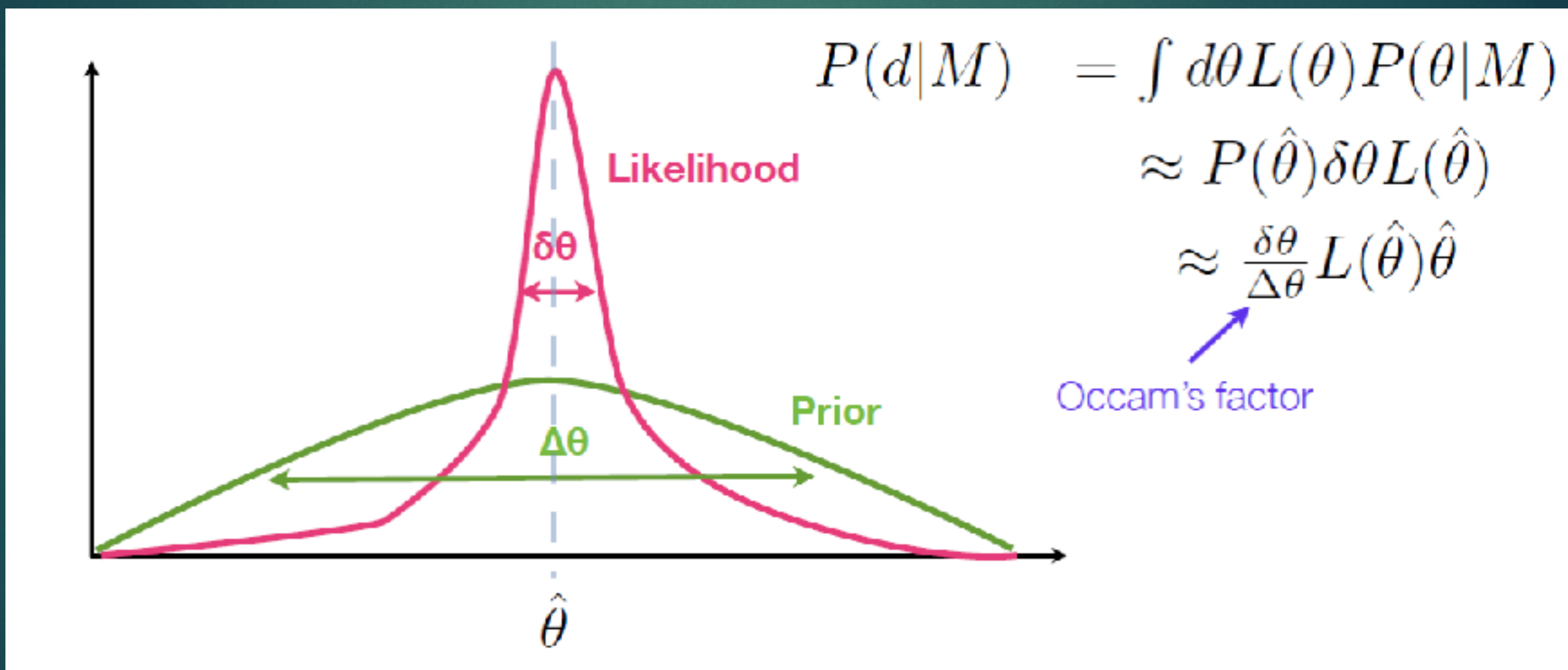
Used to evaluate the relative probabilities of different Hypothesis

Evidence is the integral of the likelihood over the prior

$$P(d|M) = \int_\Omega d\theta P(d|\theta, M) P(\theta|M)$$

The Evidence

Automatically implements Occams Razor:
A simpler model will be preferred unless the more complex one describes the data much better



The Evidence

Used to evaluate the relative probabilities of different Hypothesis

Evidence is the integral of the likelihood over the prior

$$P(d|M) = \int_\Omega d\theta P(d|\theta, M) P(\theta|M)$$

Define a 'Bayes Factor'

$$B_{01} \equiv \frac{P(d|M_0)}{P(d|M_1)}$$

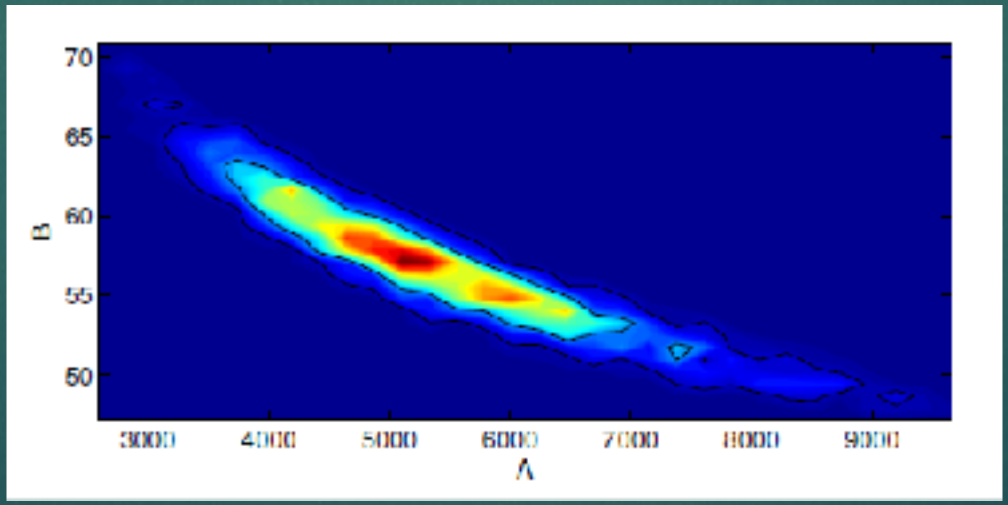
The Evidence

|lnB|	relative odds	favoured model's probability	Interpretation
< 1.0	< 3:1	< 0.750	not worth mentioning
< 2.5	< 12:1	0.923	weak
< 5.0	< 150:1	0.993	moderate
> 5.0	> 150:1	> 0.993	strong

‘Marginalisation’

Key to Bayesian analysis: Integrate over ‘nuisance’ parameters:
Things you don’t care about but that affect the answer you want to get.

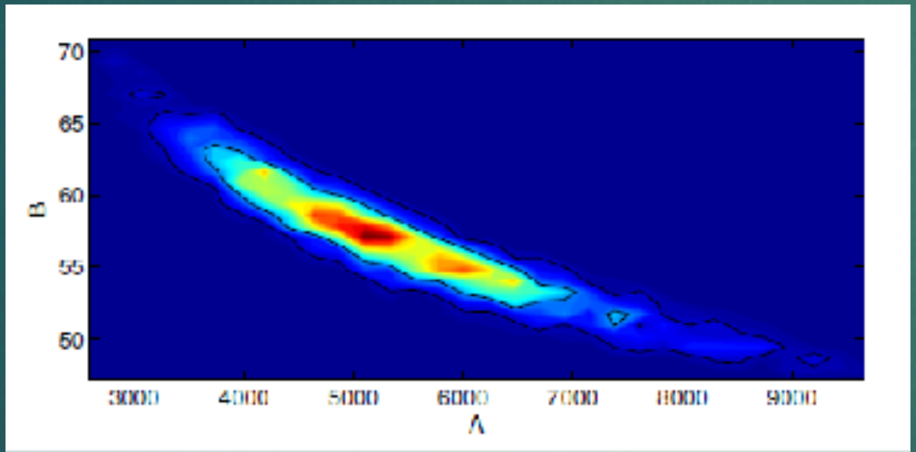
- Consider 2d problem –
- Probability density for parameters A and B.

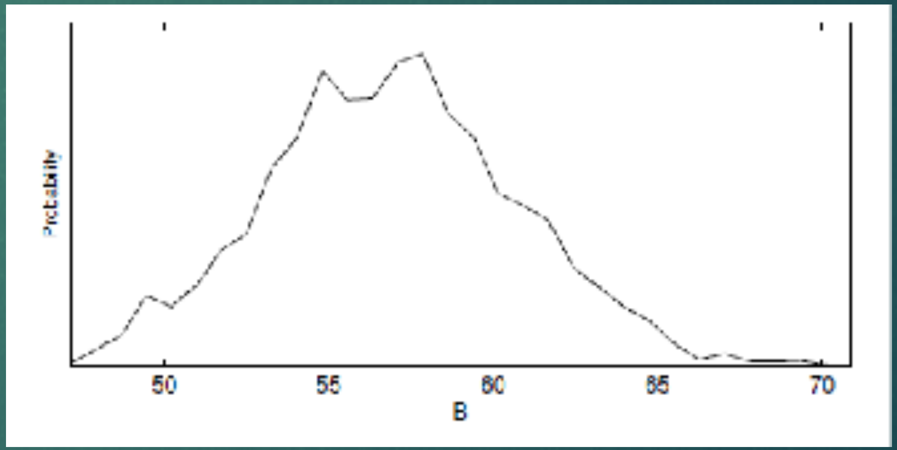


‘Marginalisation’

- Can marginalise numerically after sampling

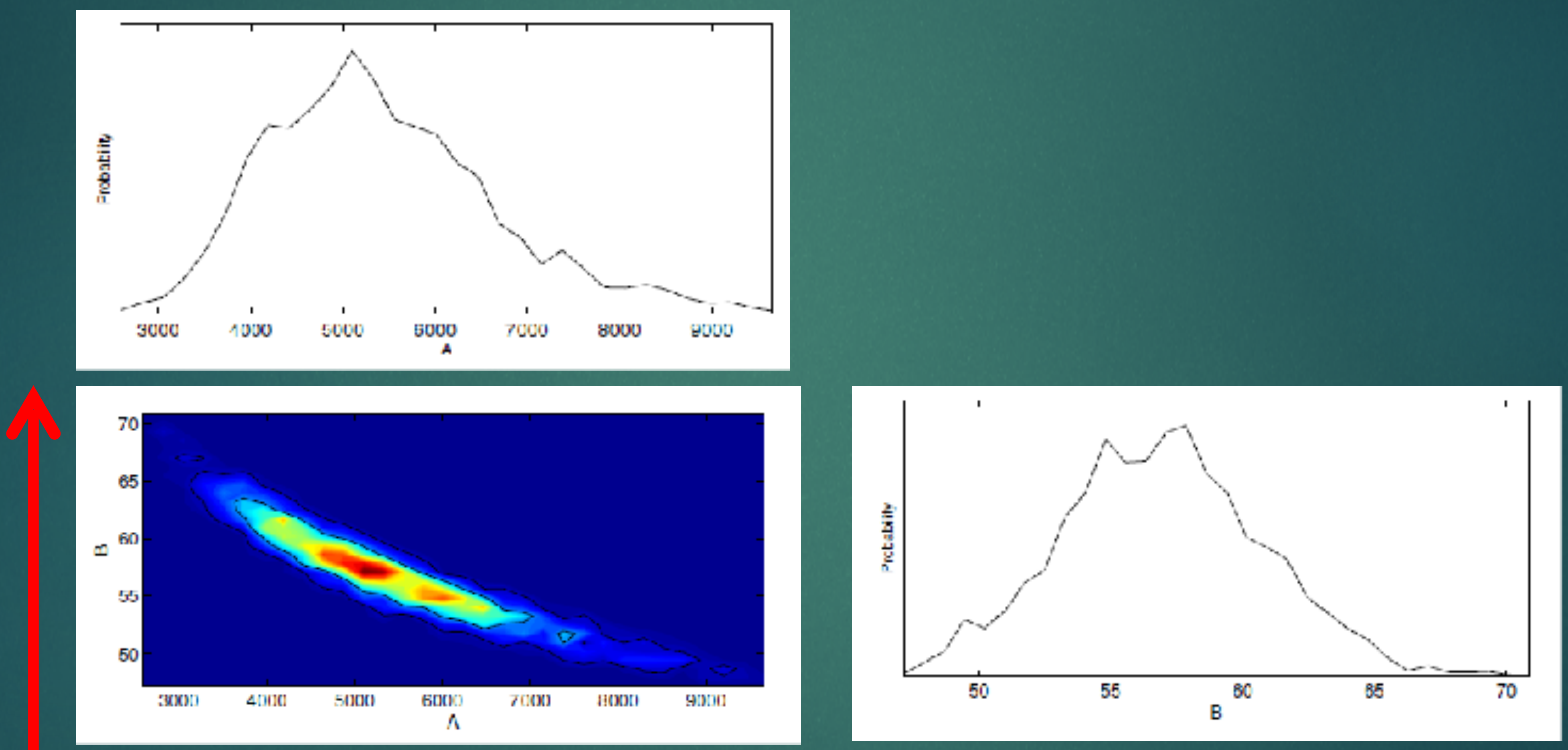
Integrate over A to get the probability of B





‘Marginalisation’

- Can marginalise numerically after sampling

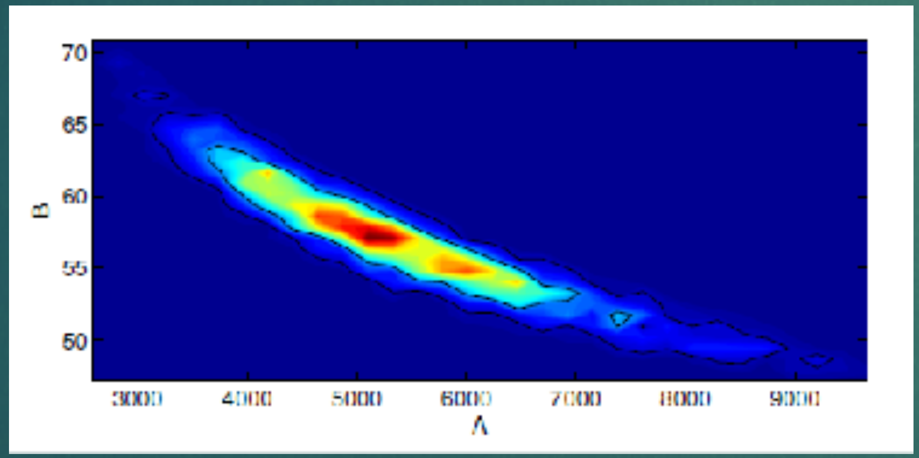


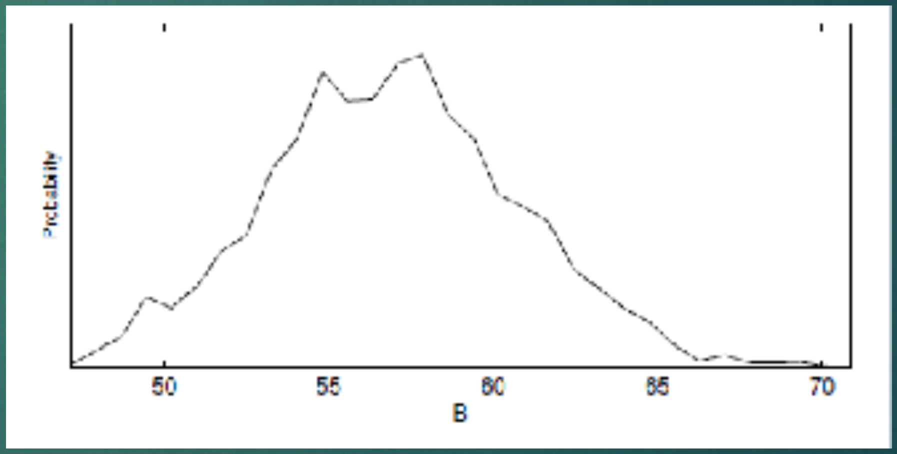
Integrate over B to get the probability of A

‘Marginalisation’

- Can also marginalise analytically

$$p(\vec{d}\,|B) = \int p(\vec{d}|A, B)p(A)dA$$



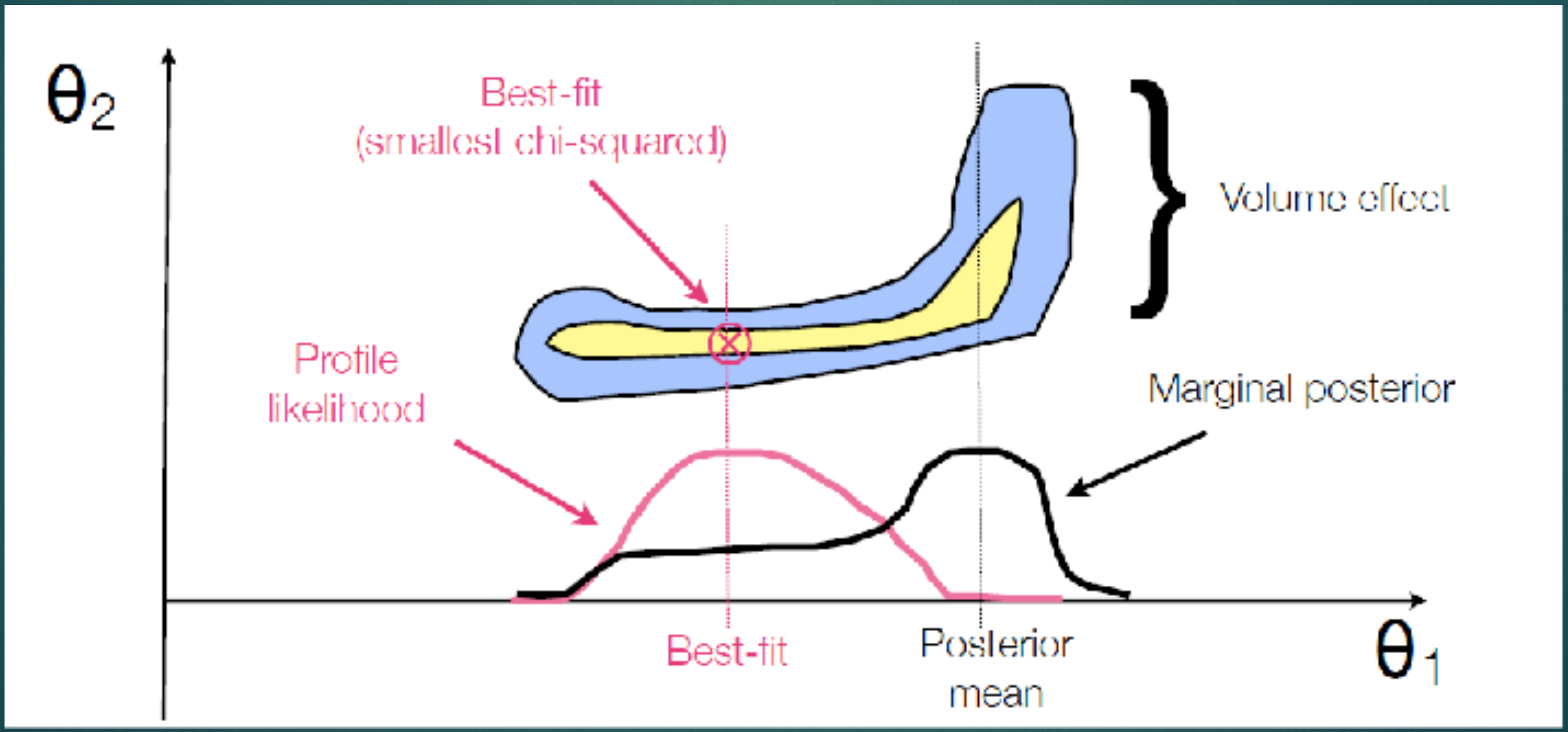


‘Marginalisation’

For uniform priors: P(M | D) = P(D | M).

Doesn’t mean Frequentist and Bayesian results will agree.

Volume Matters



Sampling

Said we want to calculate $P(X \mid D, M)$

Non-trivial for non-trivial problems

Have to sample from posterior

Markov-Chain Monte-Carlo

Markov chain – sequence of state changes that depends only on the most recent states, not the states that preceded them.

Simple example (from Wikipedia)

Probability of the weather.

Markov-Chain Monte-Carlo

P(Tomorrow is Sunny | Today is rainy) = 0.5
P(Tomorrow is rainy | Today is rainy) = 0.5

P(Tomorrow is rainy | Today is sunny) = 0.1
P(Tomorrow is Sunny | Today is sunny) = 0.9

Markov-Chain Monte-Carlo

P(Tomorrow is Sunny | Today is rainy) = 0.5
P(Tomorrow is rainy | Today is rainy) = 0.5

P(Tomorrow is rainy | Today is sunny) = 0.1
P(Tomorrow is Sunny | Today is sunny) = 0.9

P(Sun in 2 days| Sun) = P(S,S|S) + P(S,R | S) = 0.86
P(Sun in 30 days | Sun) = …………………………= 0.833
P(Sun in 100 days | Sun) = ………………………= 0.833
P(Sun in 100 days | rain) = ………………………= 0.833

Markov-Chain Monte-Carlo

Probability of weather tomorrow depends only on the last few days.

Forgets about everything previous.

Important aspect of all samplers.

It means that eventually we will always converge on the equilibrium probability no matter our starting point.

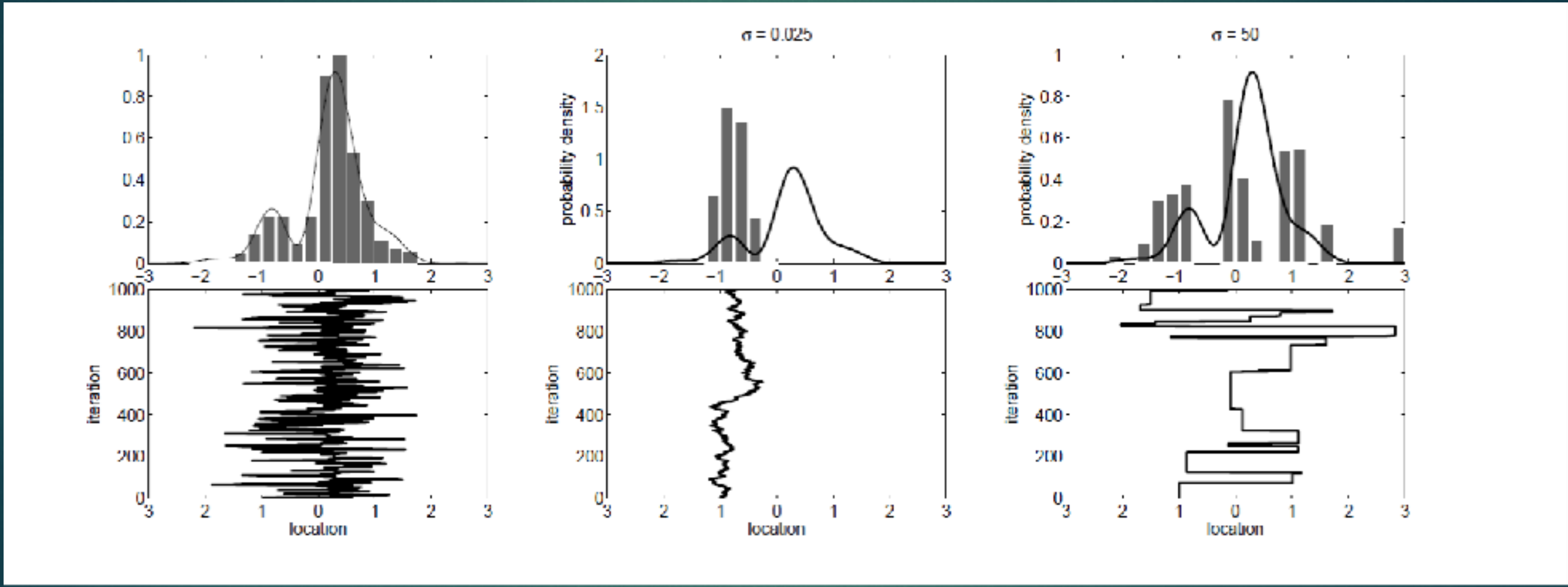
Random walk Metropolis Hastings

Simplest sampler you can imagine
~ 6 lines of Code:

Choose parameter starting point θ_0
Calculate likelihood L_0
Do:
Take a step to $\boldsymbol{\theta_1}$
Calculate likelihood L_1
Draw a random uniform number U from 0..1
If $L_1/L_0 > U$ accept the new point, otherwise reject.
Repeat.

Random walk Metropolis Hastings

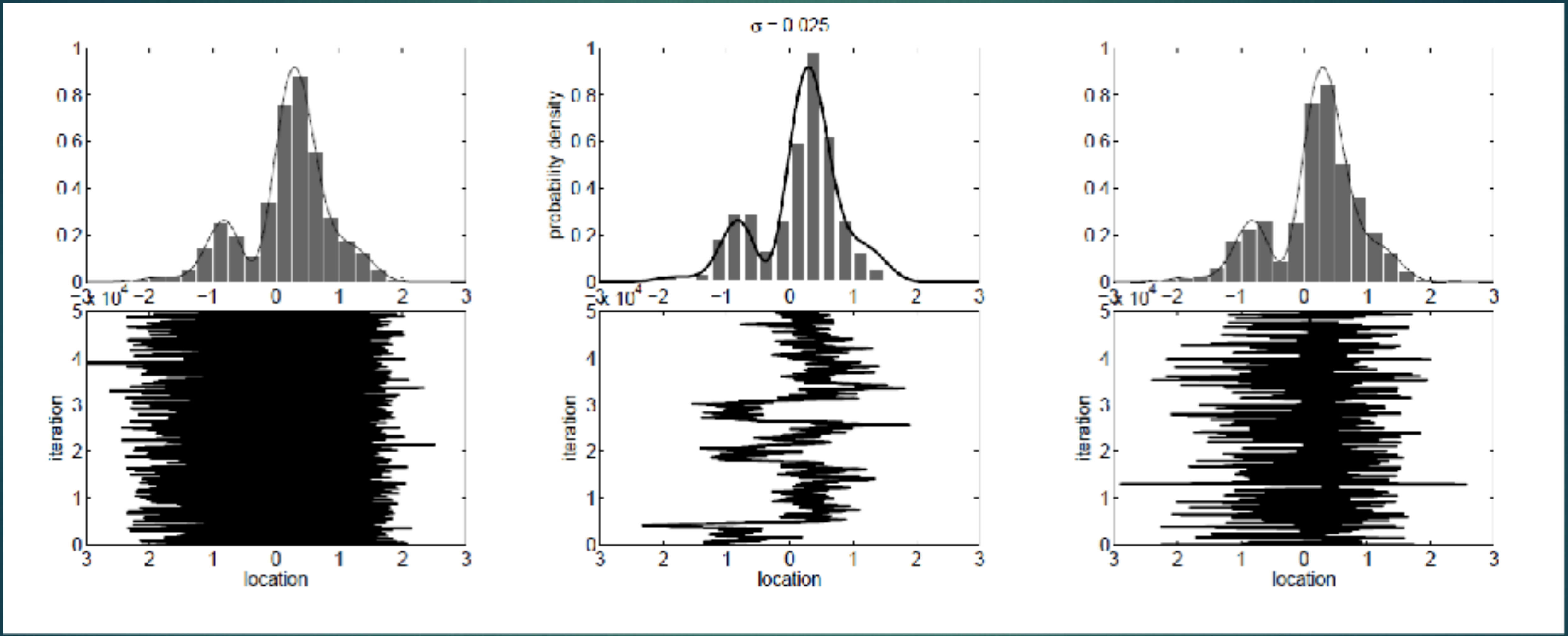
Has its problems: Convergence rate depends on step size



Step Size: Just right Too small Too big

Random walk Metropolis Hastings

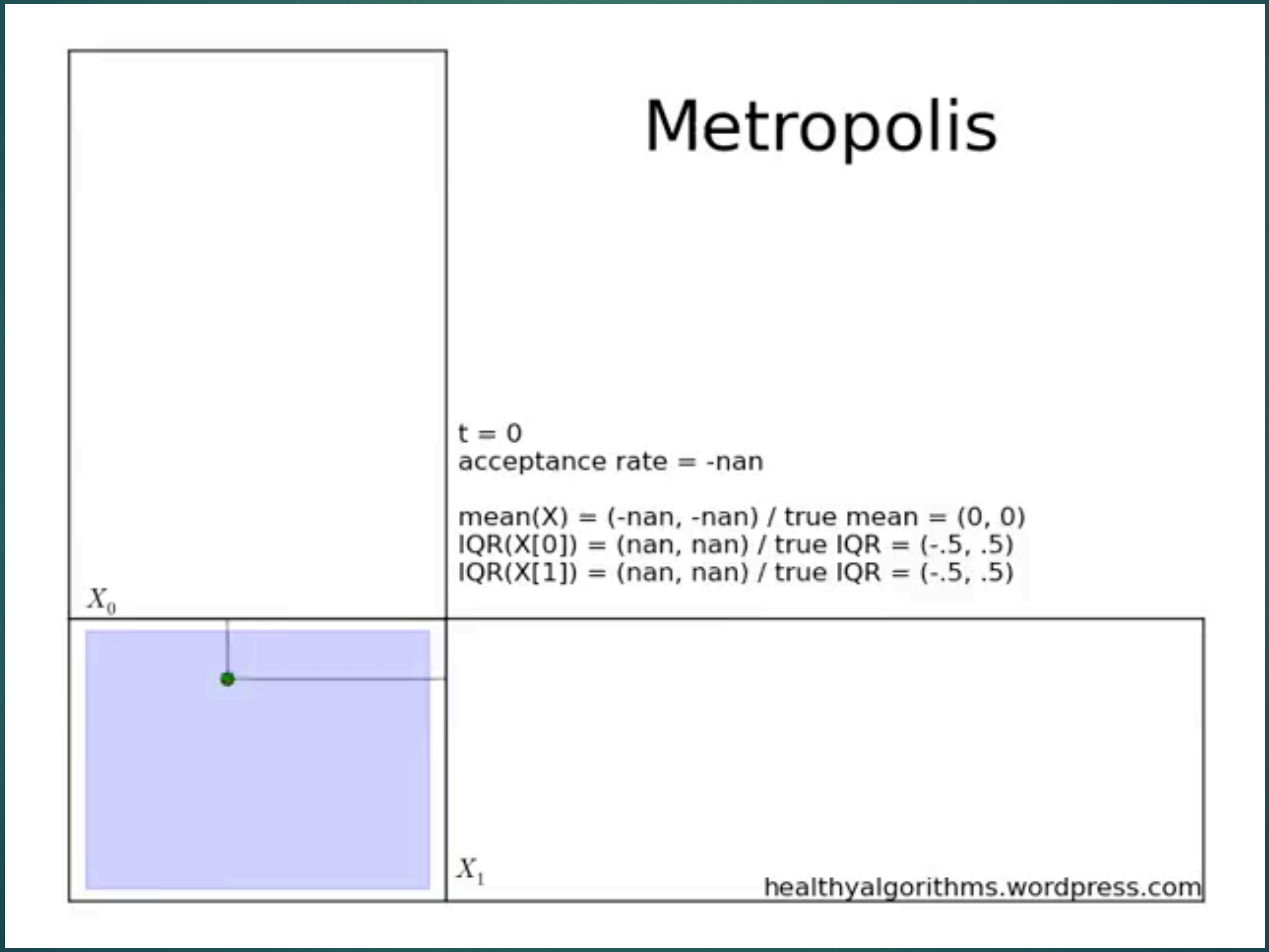
But will get there eventually



Step Size: Just right Too small Too big

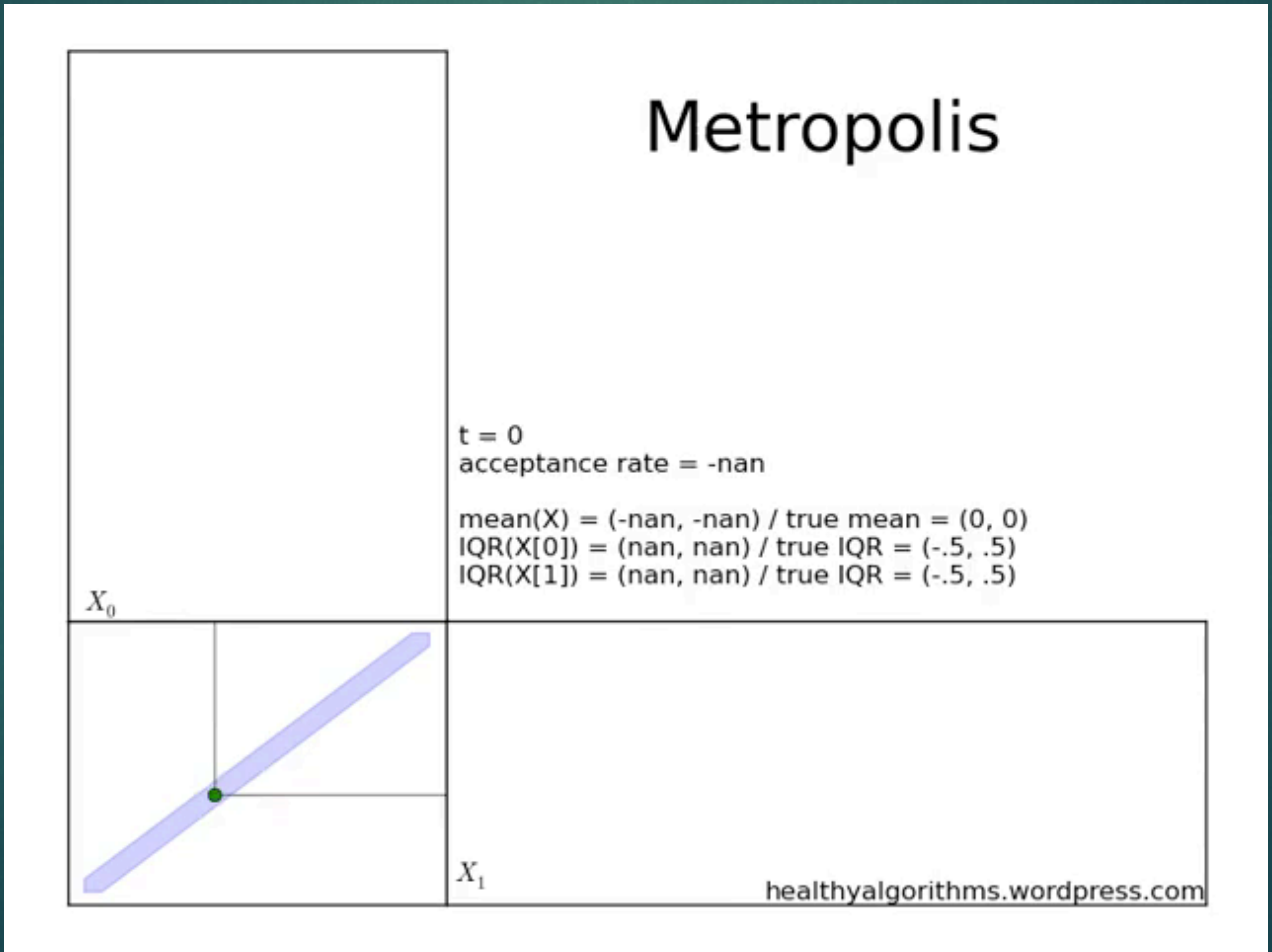
Random walk Metropolis Hastings

For simple problems though it is all you need.
E.g. Unit Square:



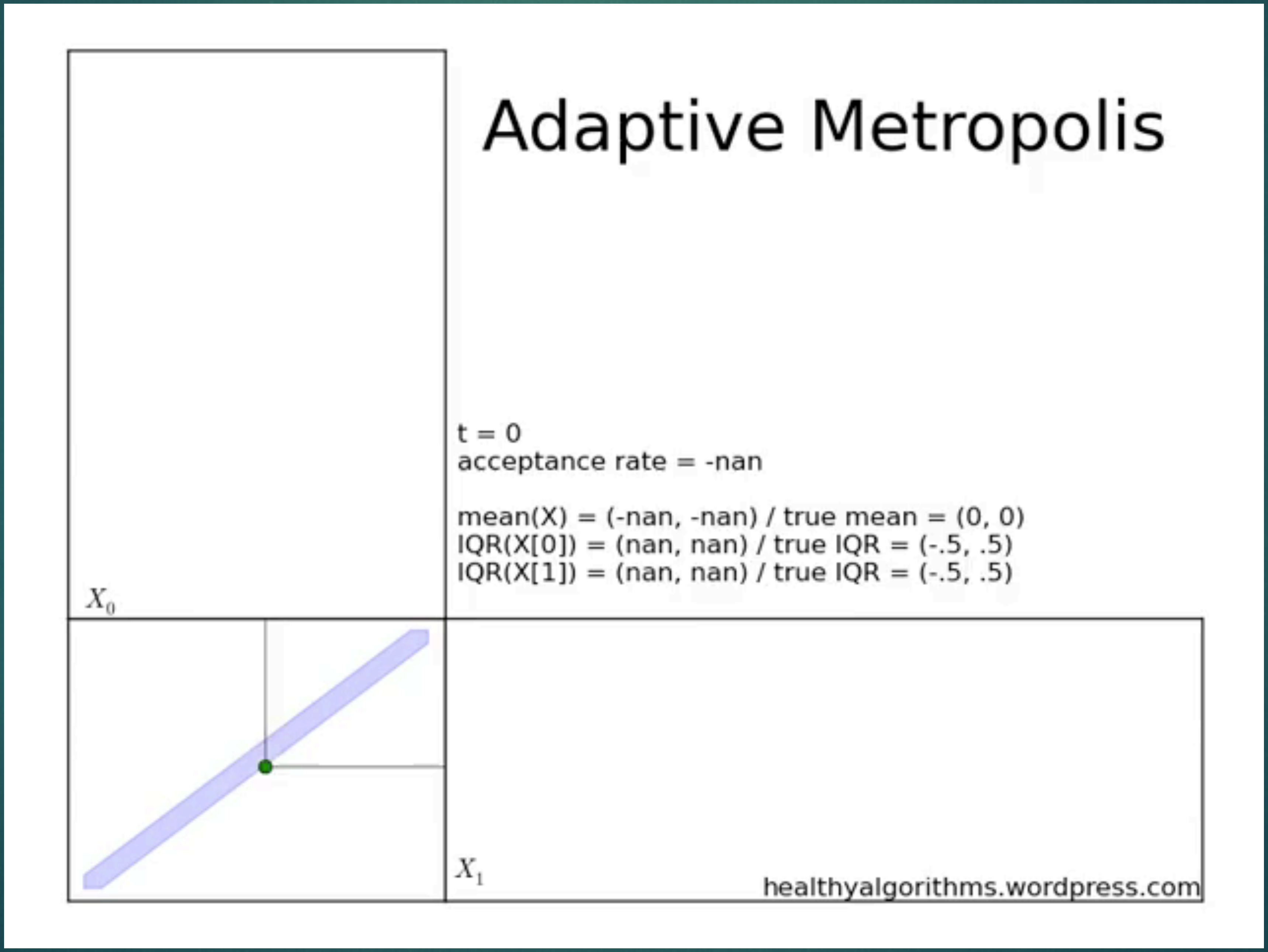
Random walk Metropolis Hastings

Quickly becomes insufficient for more complex problems: 2D covariant parameters



Random walk Metropolis Hastings

Adaptive Metropolis much better solution.
Adapts step size to decrease autocorrelation length.



Metropolis Hastings

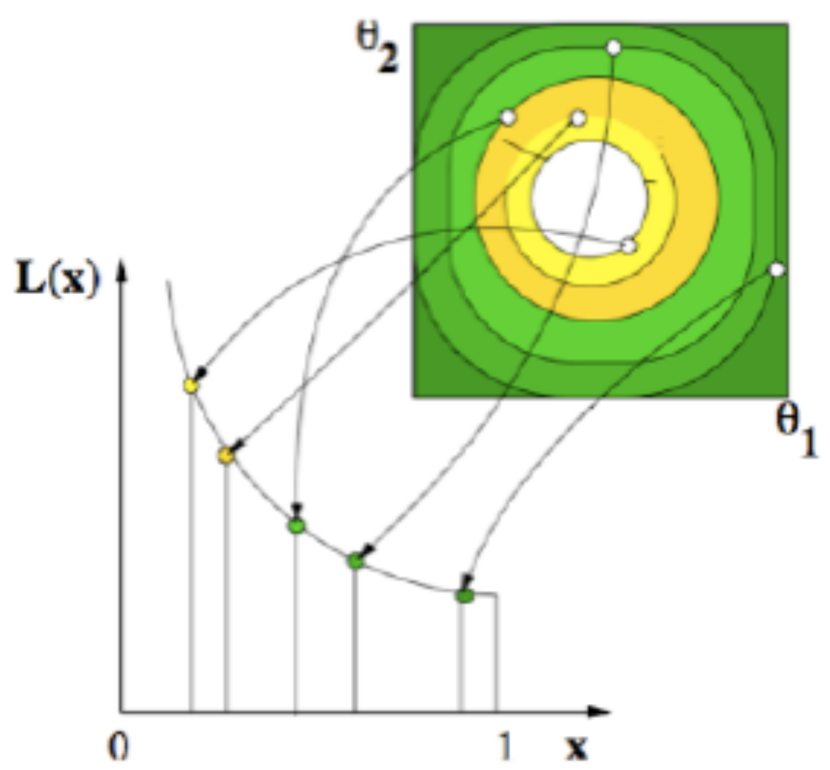
Generally very poor for multi modal problems:

If step size allows jumps between modes,
it will be too big within each mode.

If step size small enough to explore individual modes,
it wont step between them.

Nested Sampling (Skilling 2004)

Solves a lot of these problems



Liddle et al (2006)

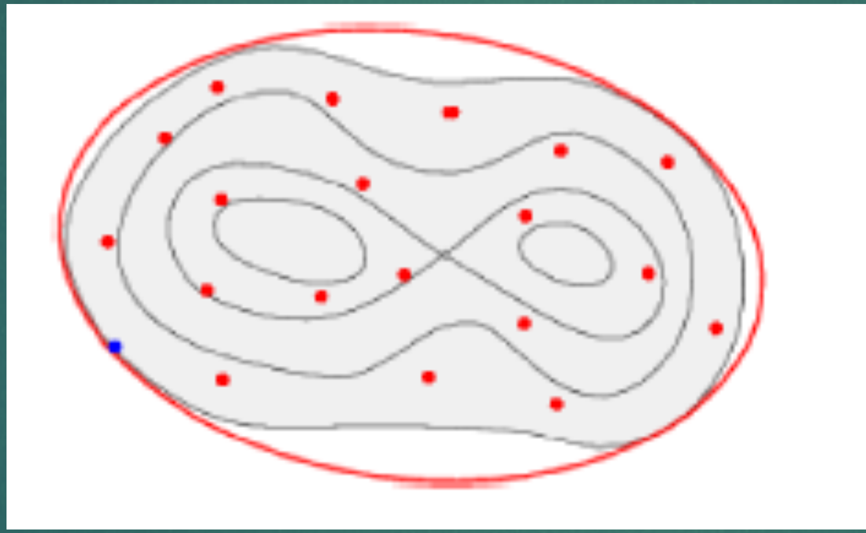
Draw N points Uniformly from the prior
Lowest likelihood point = L_0

Draw a new point with likelihood L_i
If $L_i > L_0$ replace point with the new point

Otherwise try again

Nested Sampling (Skilling 2004)

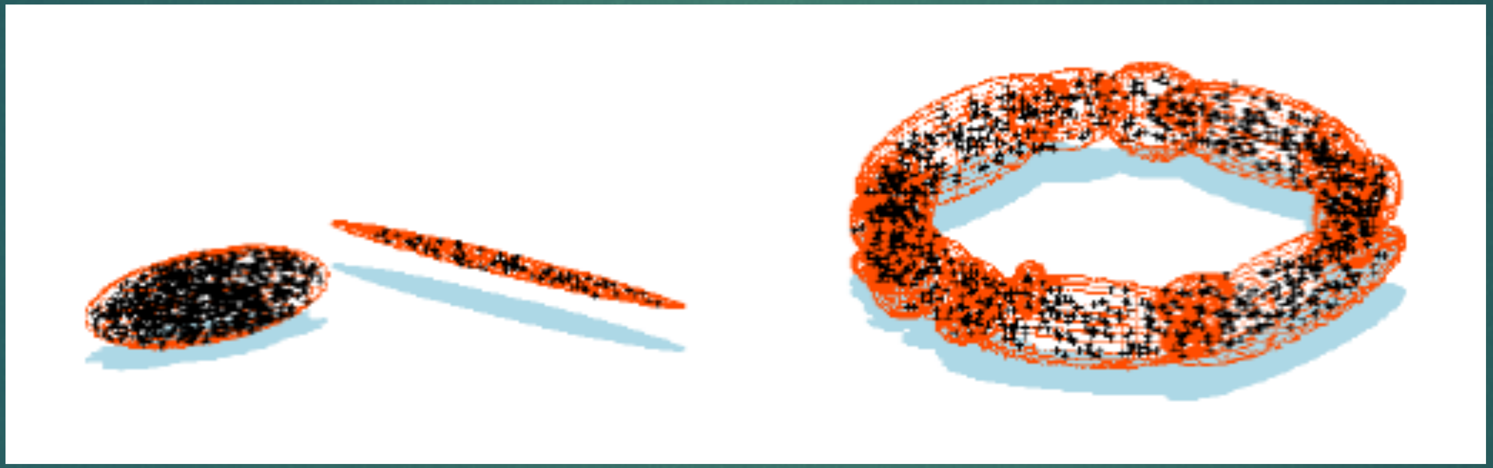
The Challenge:
Draw new points from within the hard boundary $L > L_0$
Mukherjee (2005): Use ellipses to define the boundary



Still wasn't great for multi-modal problems.

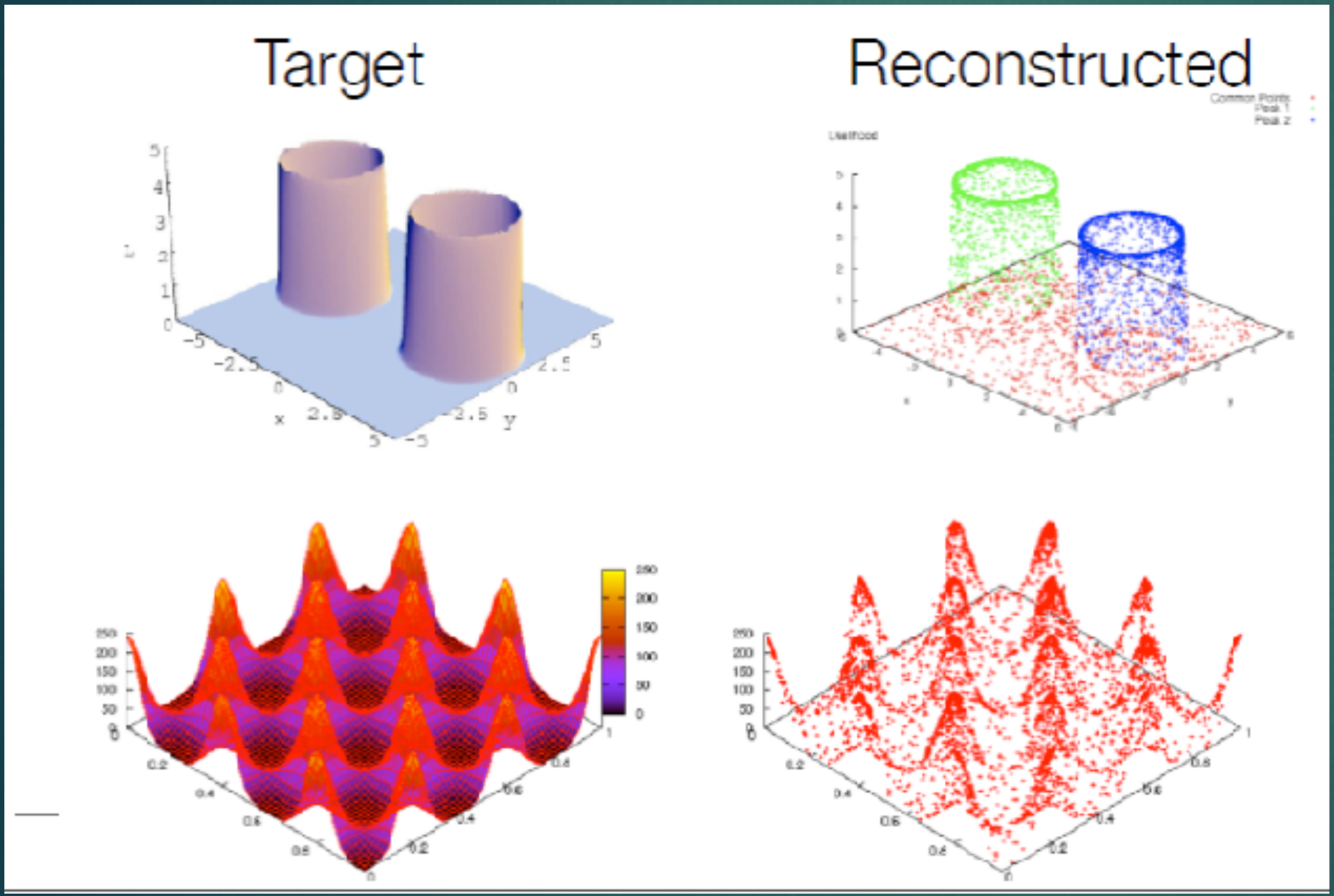
MultiNest (Feroz & Hobson 2008)

At each iteration:
Construct optimal multi-ellipsoidal bound
Pick ellipse at random to sample new point



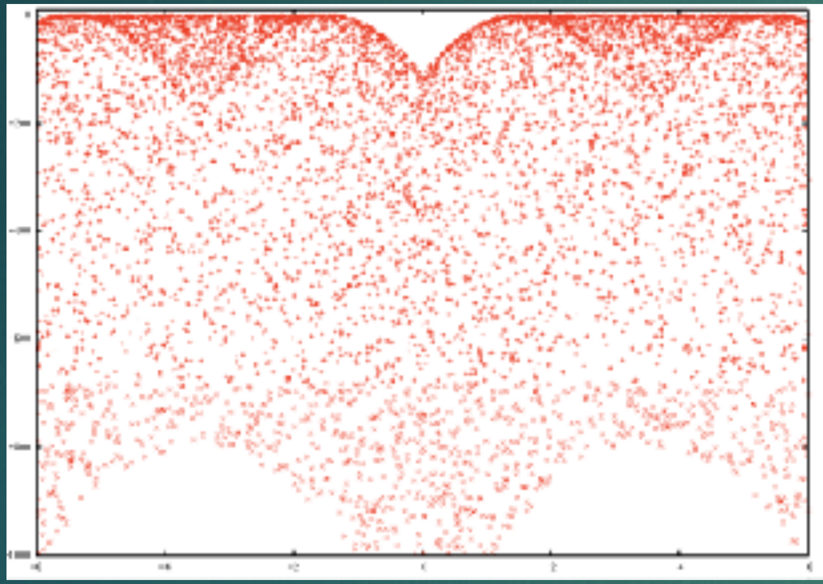
MultiNest (Feroz & Hobson 2008)

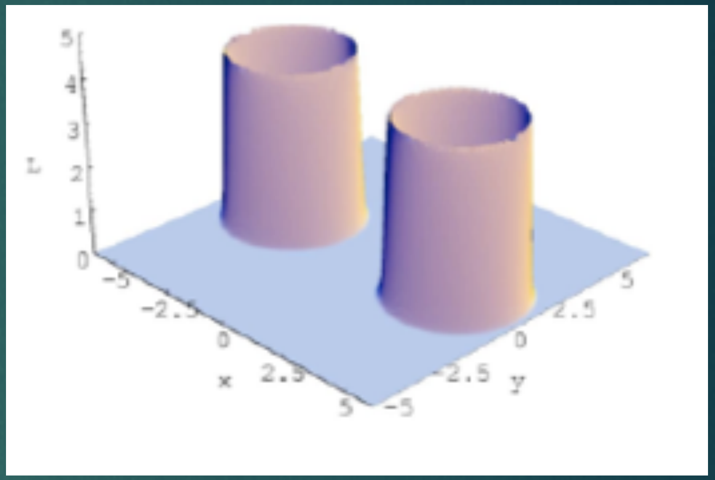
Works great for multi-modal problems:

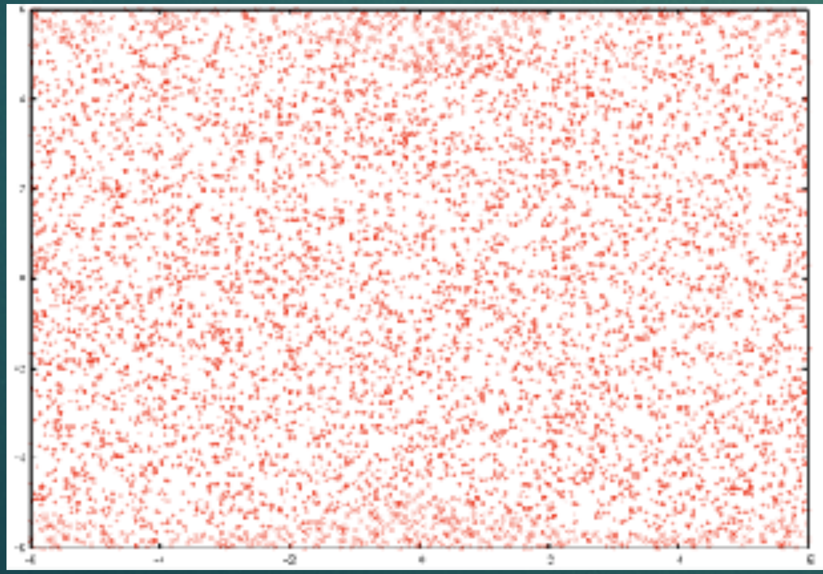


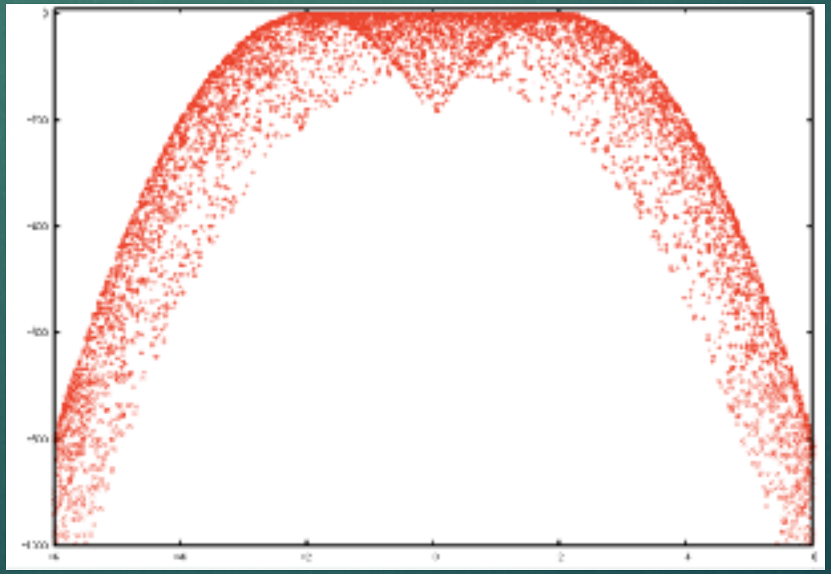
E.g. Gaussian Shells:

Start by sampling uniformly from prior in 2-dim:



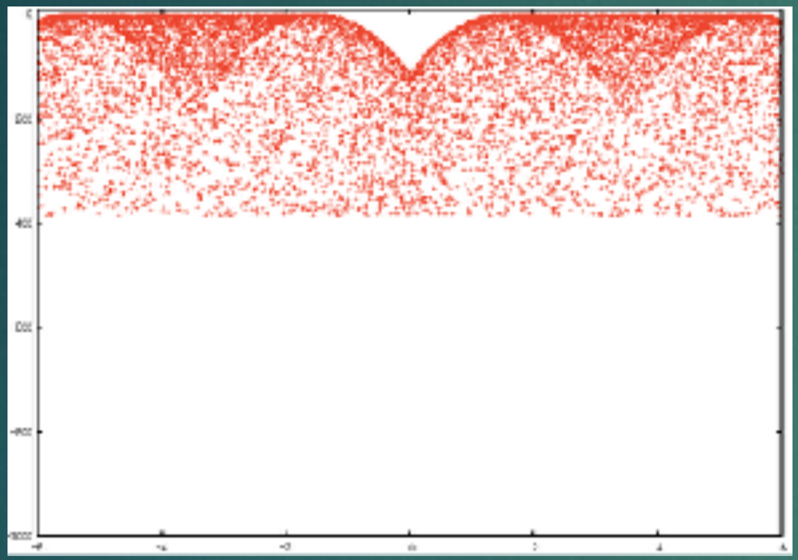


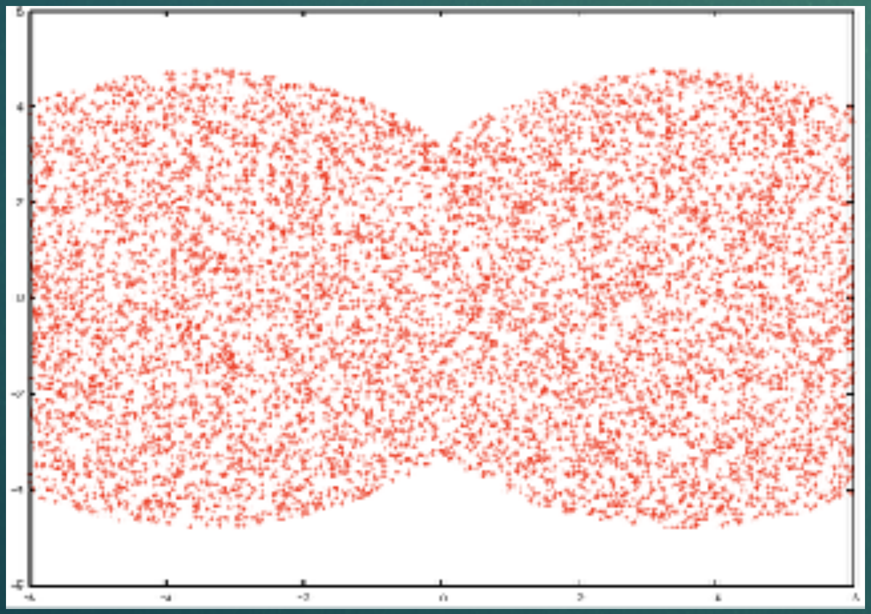


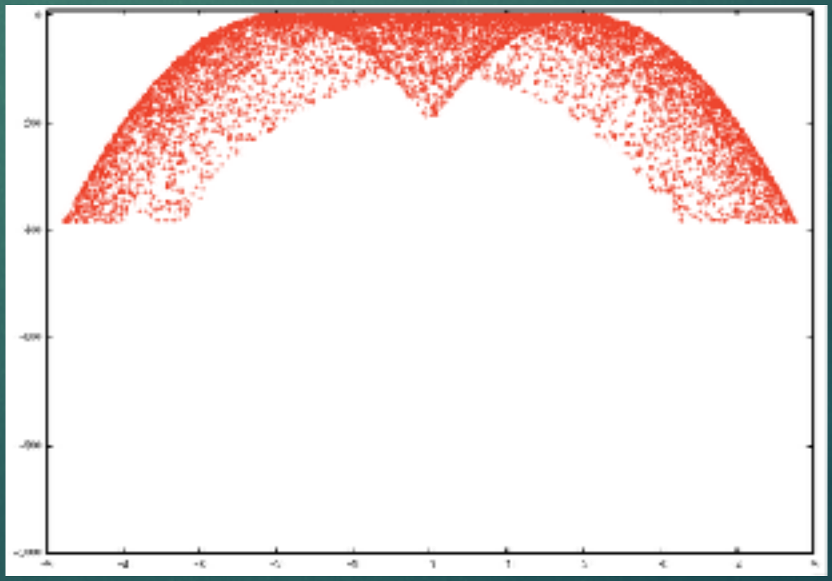


E.g. Gaussian Shells:

Then algorithm ‘nests’ upwards in likelihood

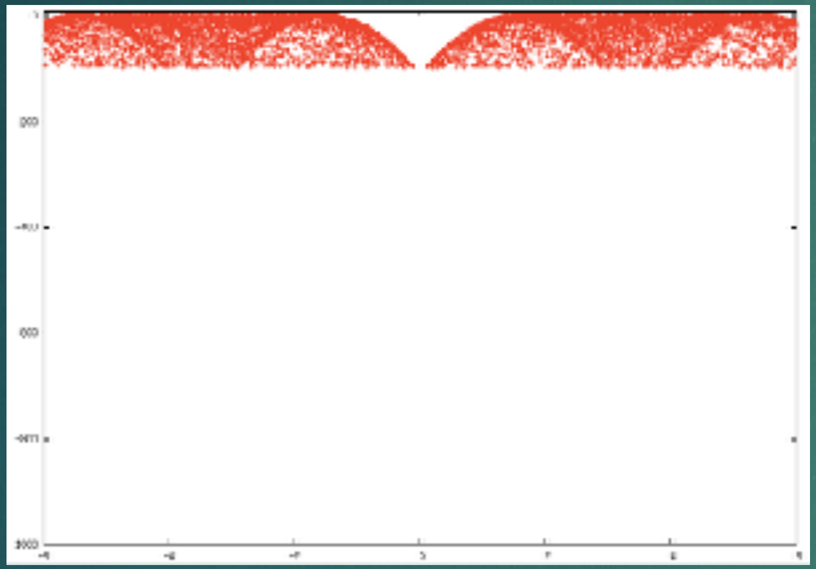


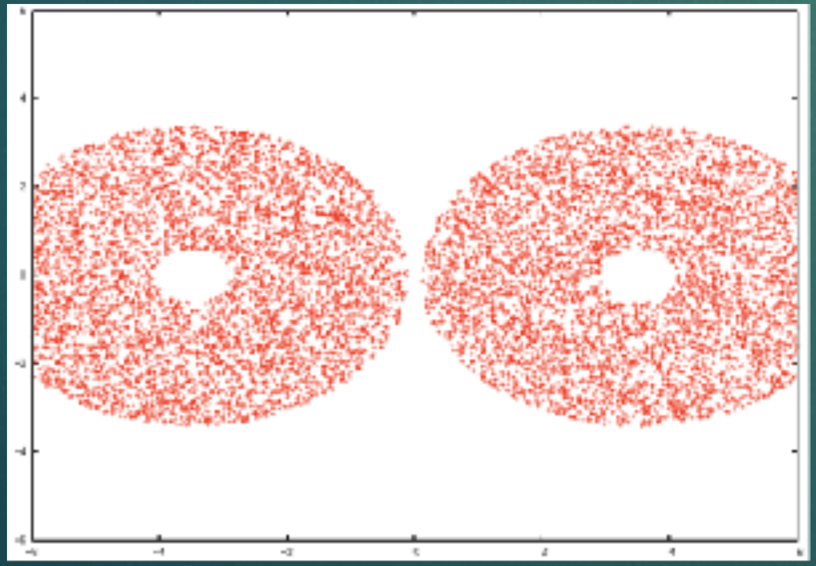


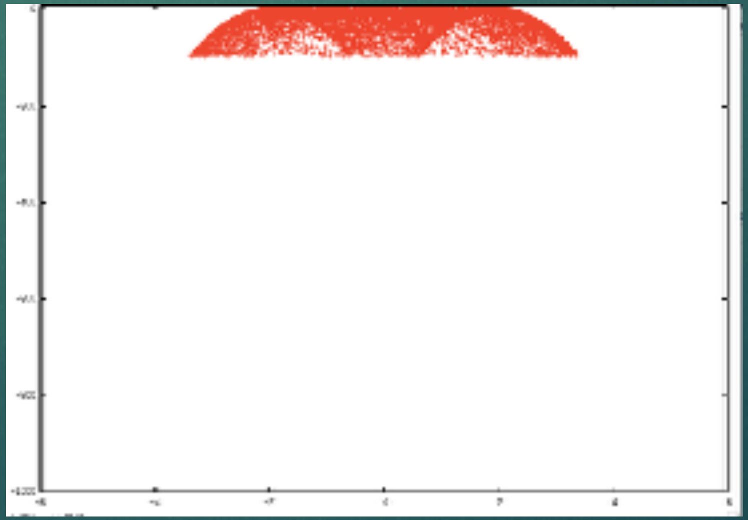


E.g. Gaussian Shells:

Then algorithm ‘nests’ upwards in likelihood

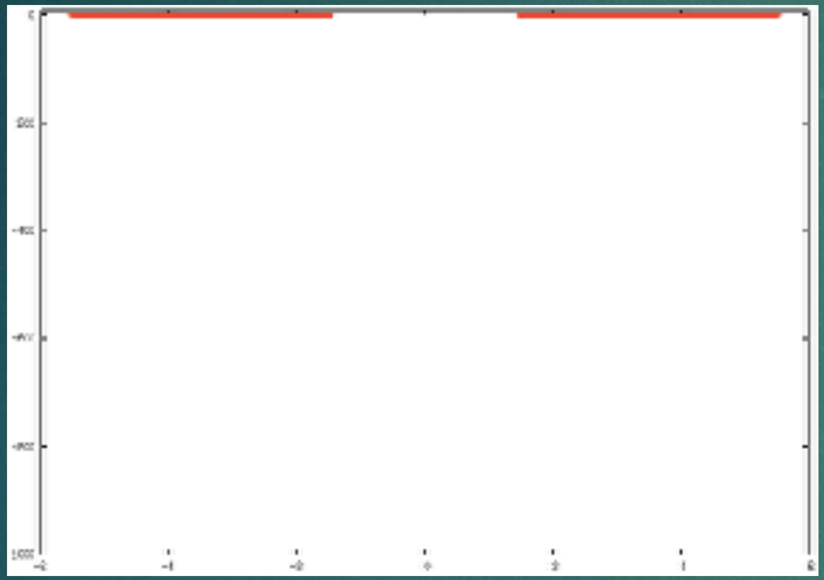


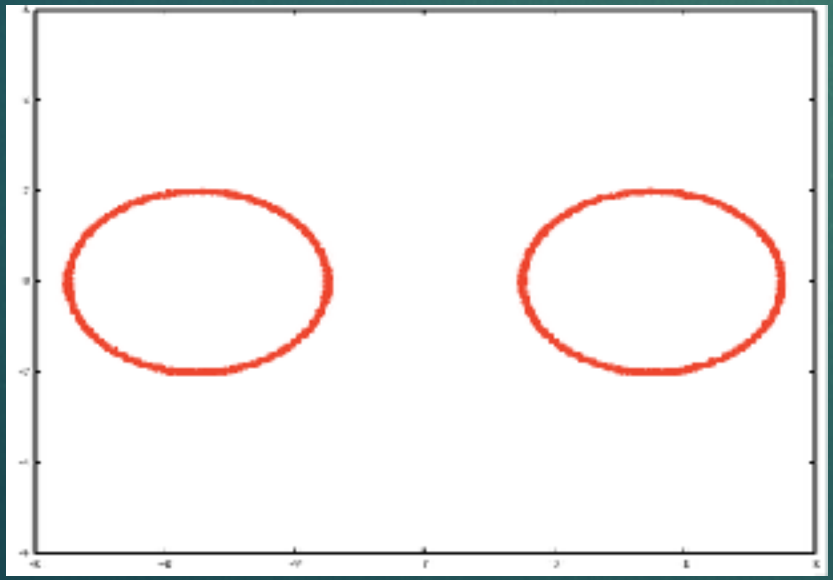


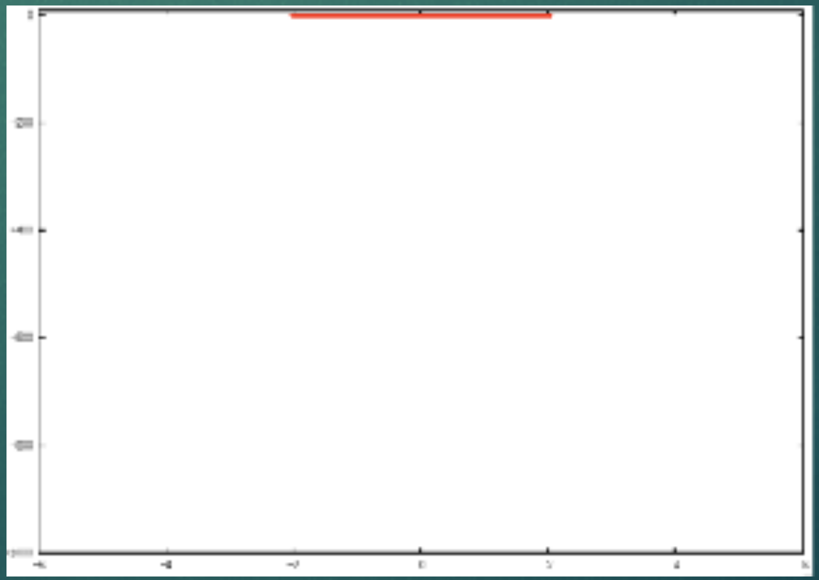


E.g. Gaussian Shells:

Then algorithm ‘nests’ upwards in likelihood

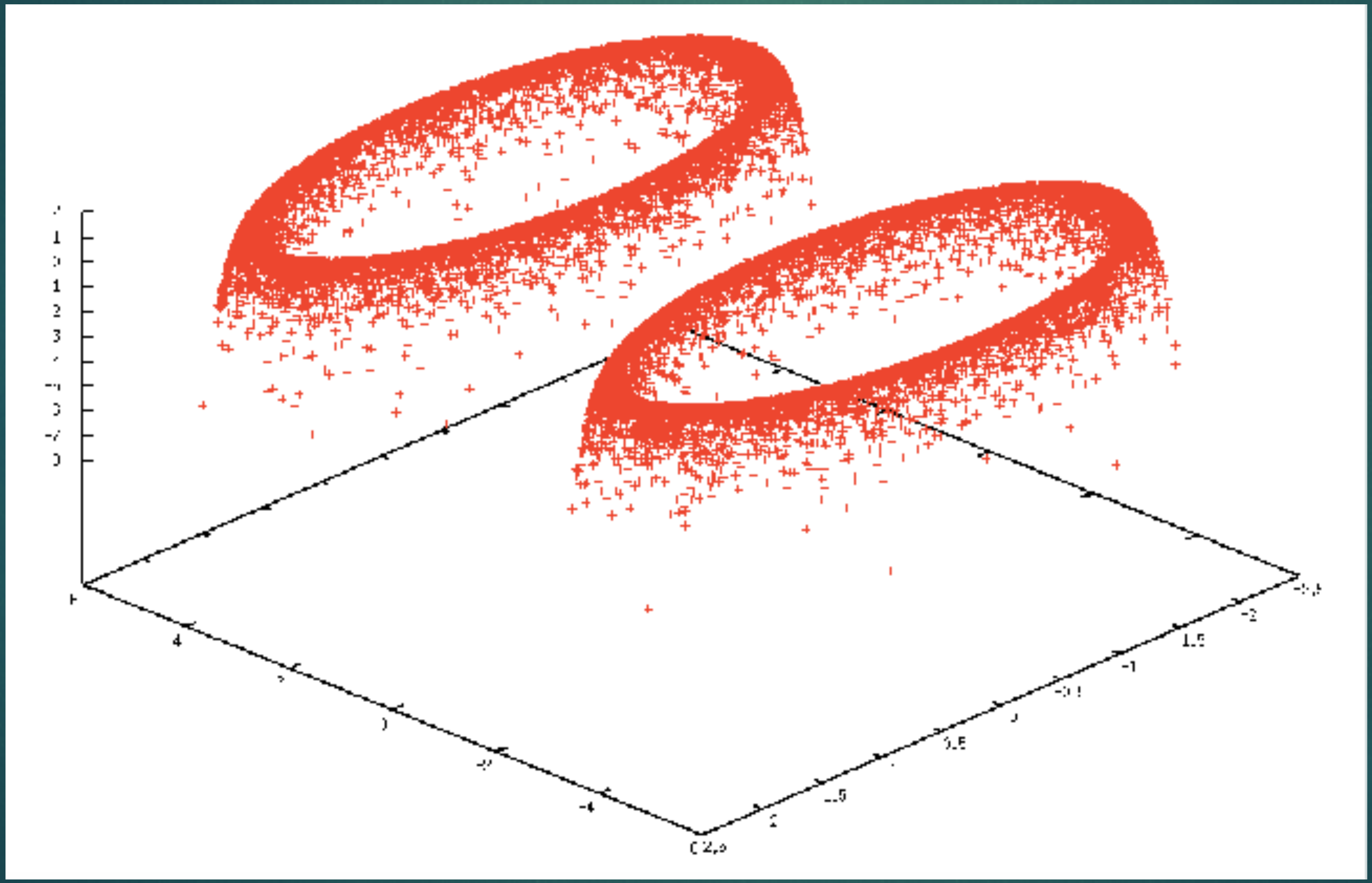






E.g. Gaussian Shells:

After sampling you have your posterior probability distributions.



Polychord (Handley & Hobson 2015)

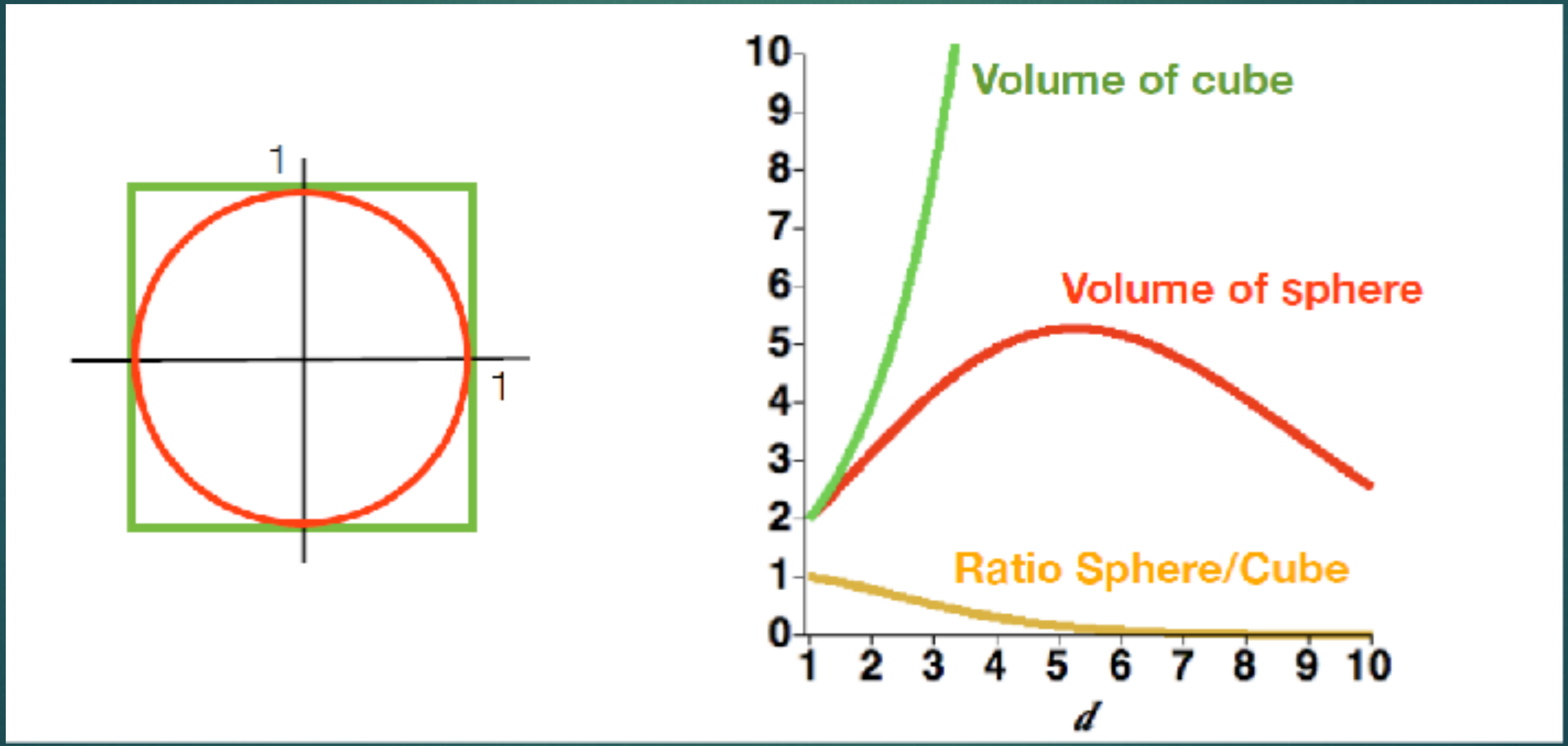
Successor to MultiNest.

Still uses nested sampling.

Works in much higher dimensions (up to ~ 150)

Nested Sampling

Dimensionality still a problem
Volume in a hypercube is dominated by the edge



Hamiltonian Monte Carlo

Very Different approach to sampling.

Able to sample millions of dimensions.

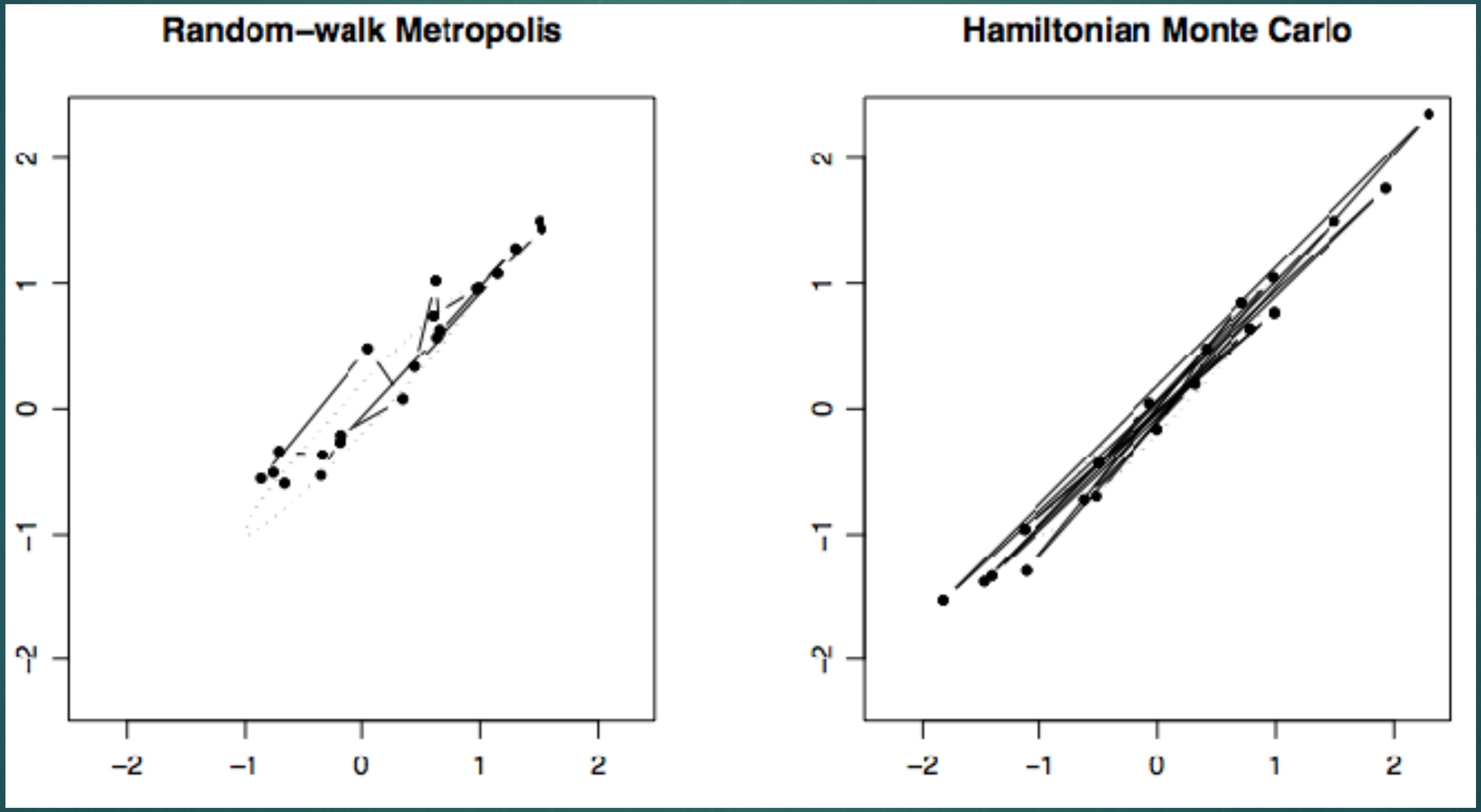
Uses gradient information to evolve the system using Hamiltonian mechanics.

Define Hamiltonian as:

$$H = \sum_{i}^{N} \frac{p_i^2}{2m_i} + \psi(\mathbf{x}).$$

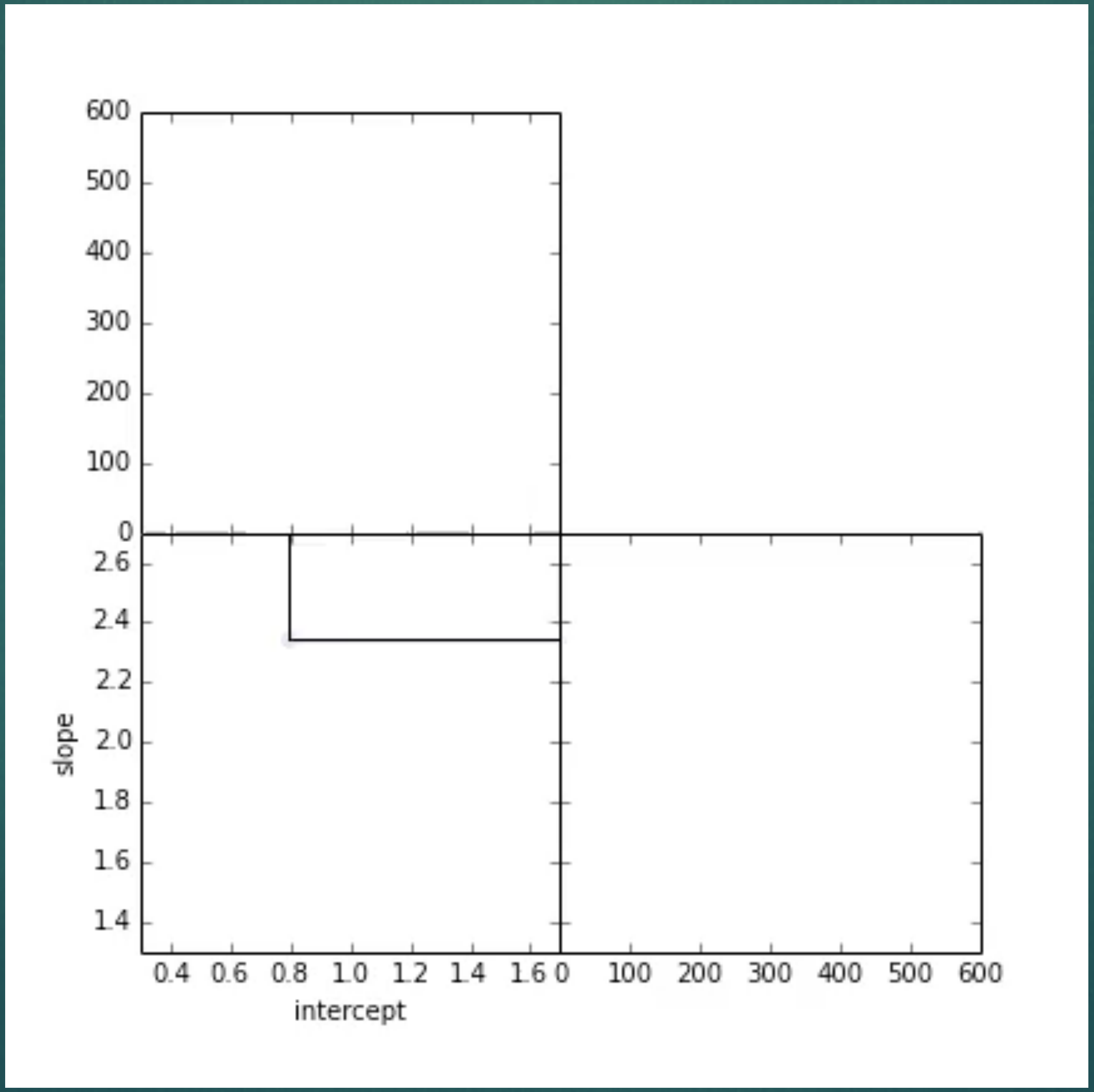
Hamiltonian Monte Carlo

More complicated – but reduces random walk



Hamiltonian Monte Carlo

More complicated – but reduces random walk



Hamiltonian Monte Carlo

Downside: Lots of tuneable parameters still (1 mass per parameter).

'Guided' Hamiltonian sampling solves this (Balan et al in prep)

Uses Hessian to define a step size matrix, accounting for correlations
In principle leaves only 1 tuneable parameter (overall step size).

Can still require 'tuning' runs if the Hessian is a poor approximation to the true likelihood.

Ideally would like some kind of adaptive hamiltonian monte carlo (anyone?)

Part 2:
Data Problems
(Or why we havn’t detected gravitational waves yet)

What can we say about pulsars?

They are very precise clocks.

Some
Pulsars are very precise clocks

This is the crab pulsar →

Radiation from the pulsar creates shocks That are felt for ~ 10 light years

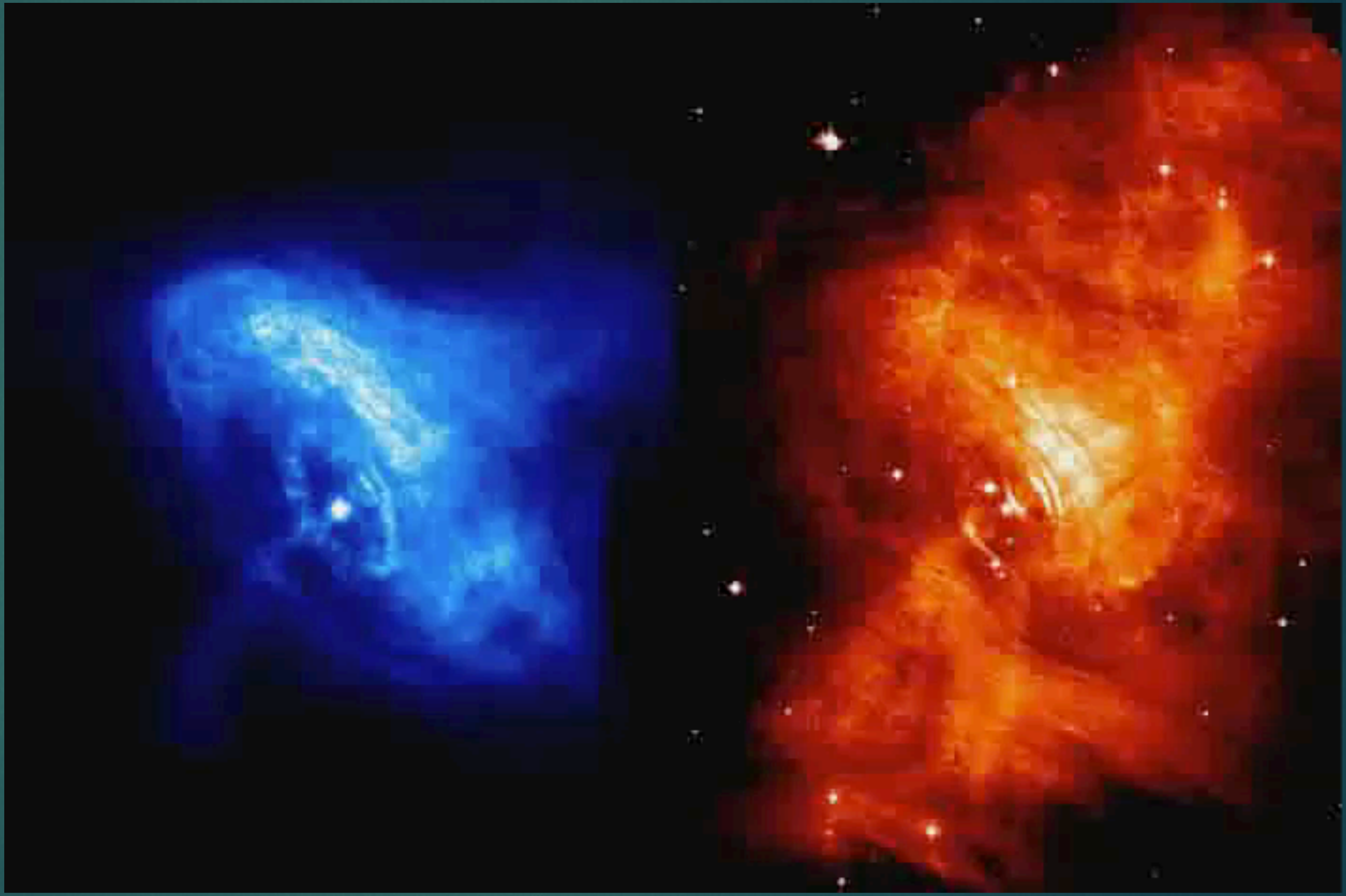


Fig: NASA

Most Pulsars are rubbish clocks

But Crab not a stable rotator:

Period of rotation has significant variation with time

No good for GW science.

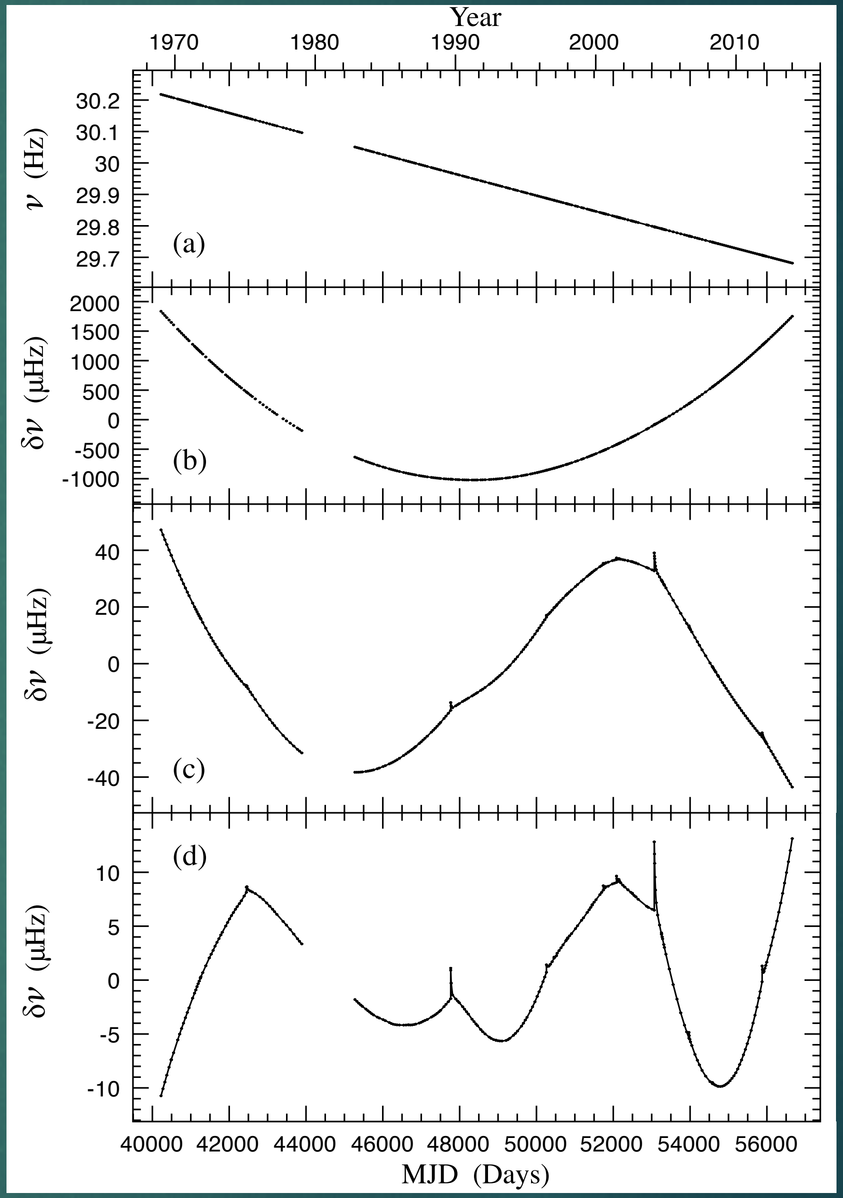
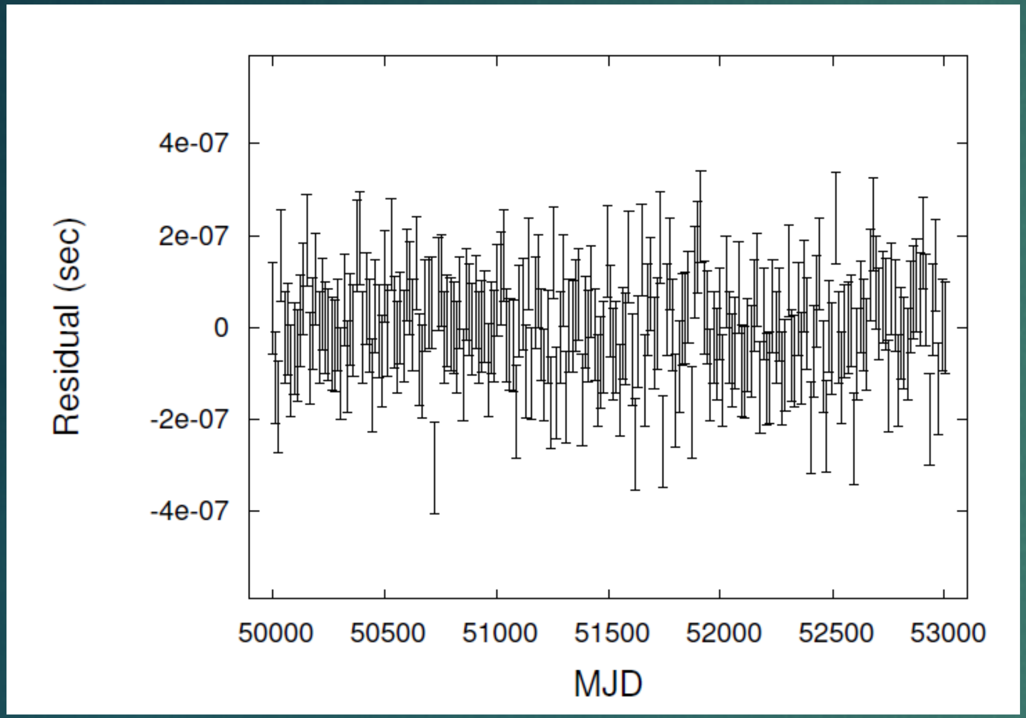
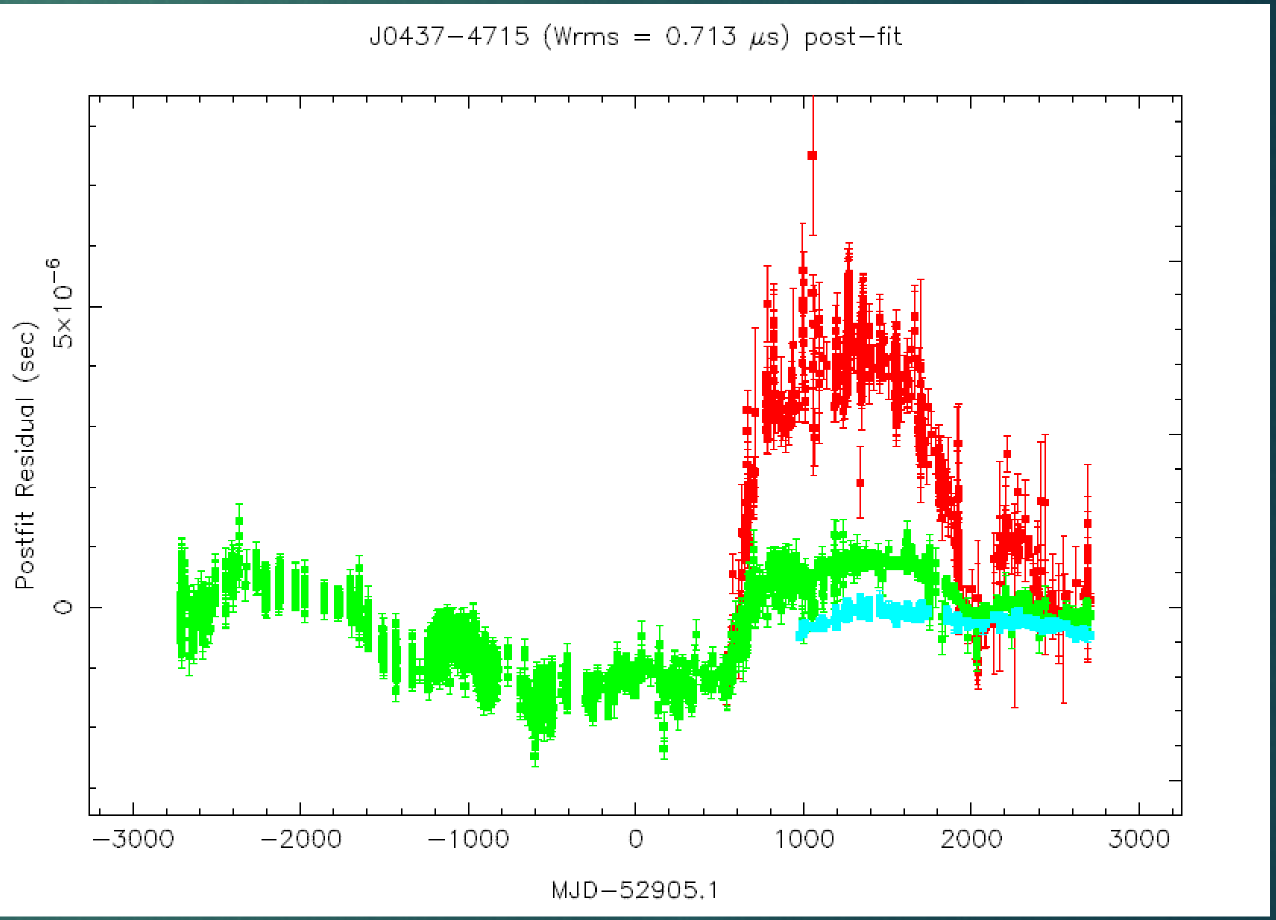


Fig: Lyne et al 2014

Data Challenges



<- 100 ns white noise
(as per early predictions)

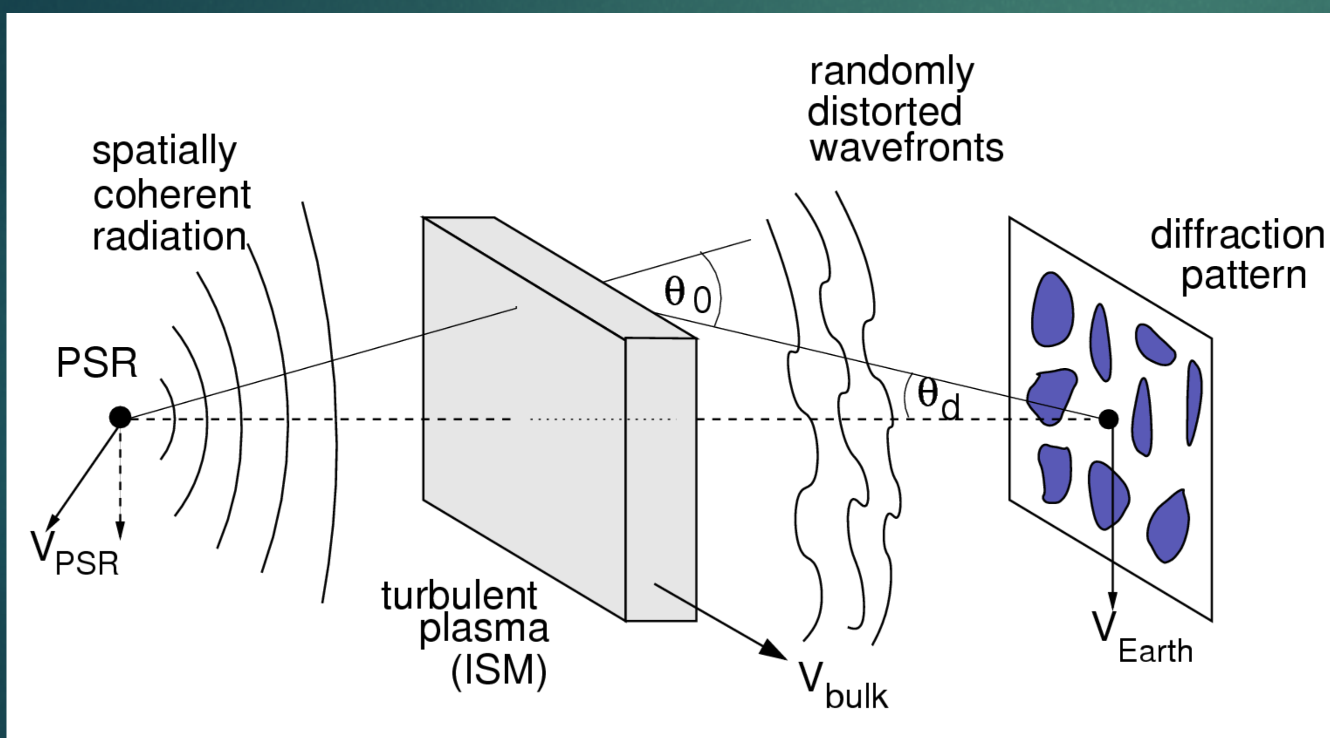


Actual Data ->

J0437-4715
(That great one mentioned earlier...)

Data Challenges

In this case noise mostly due to the interstellar medium.



Dependent on observing frequency

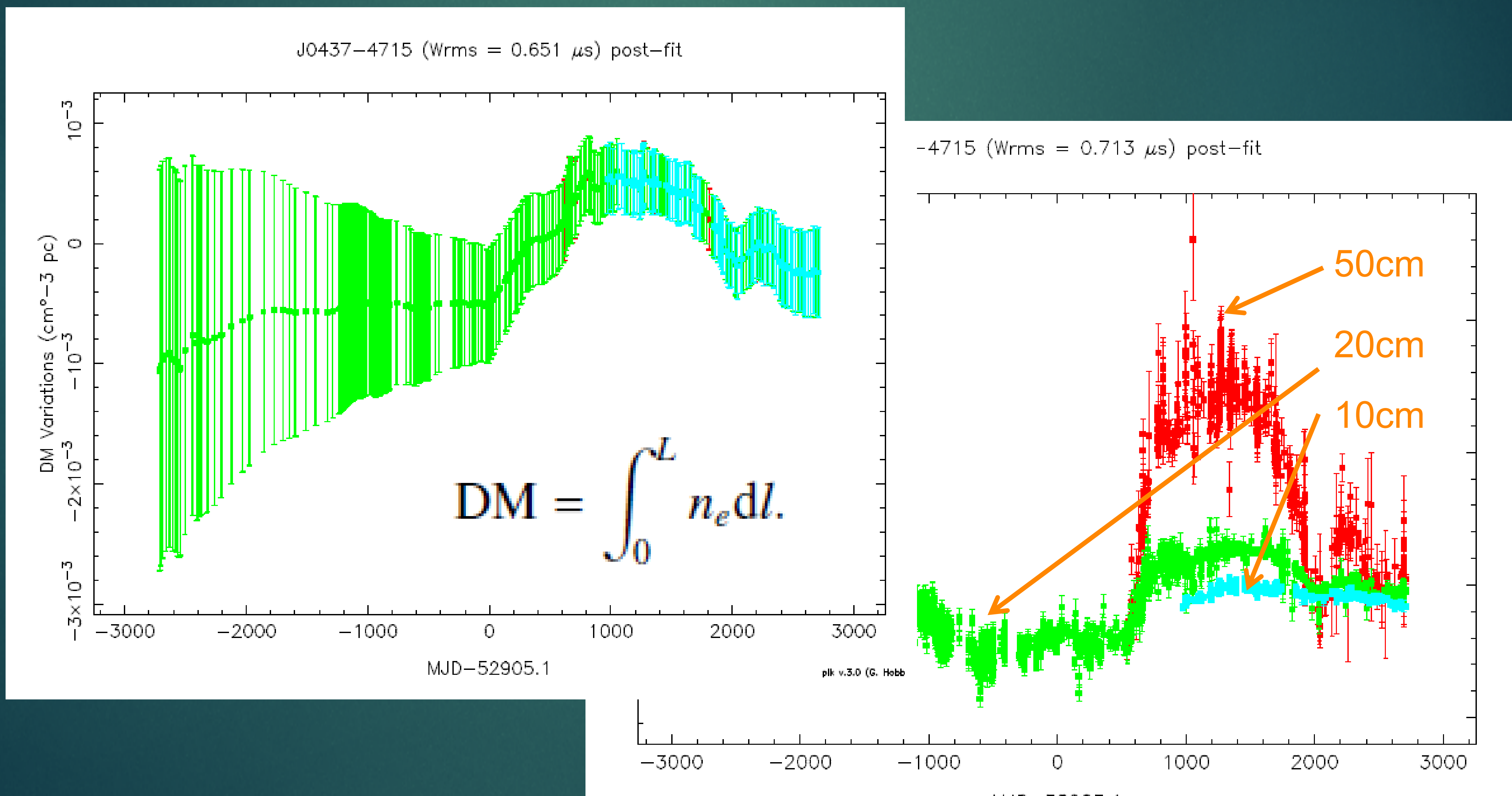
$$t_g(\nu) = K\,DM/(\nu^2)$$

$$K \equiv 4.15 \times 10^{15}\ \mathrm{Hz^2\ cm^3\ pc^{-1}\ s}$$

$$\mathrm{DM} = \int_0^L n_e \mathrm{d}l.$$

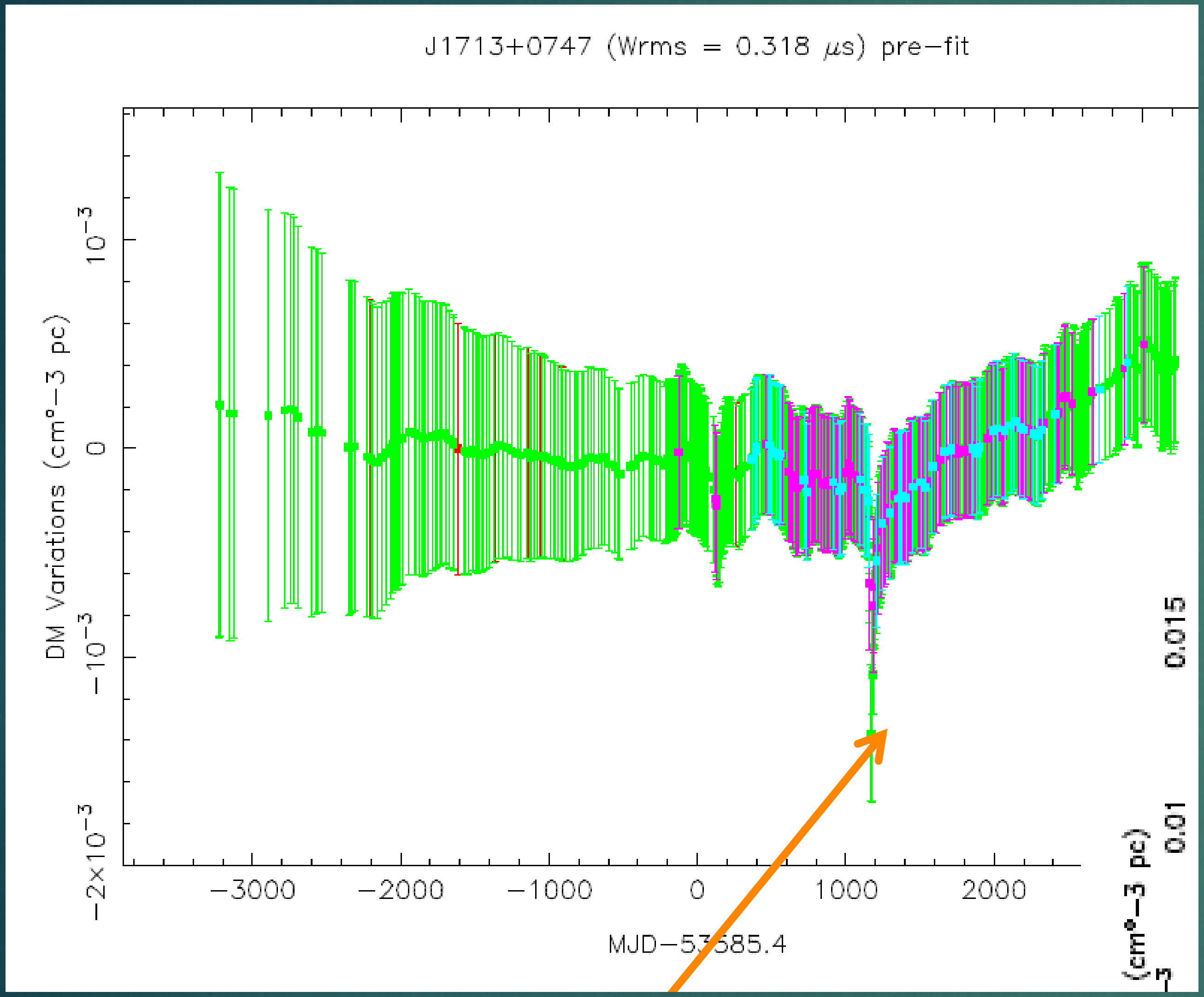
Data Challenges

Model signal statistically -
Scale with observing frequency
(You'll be doing this later)



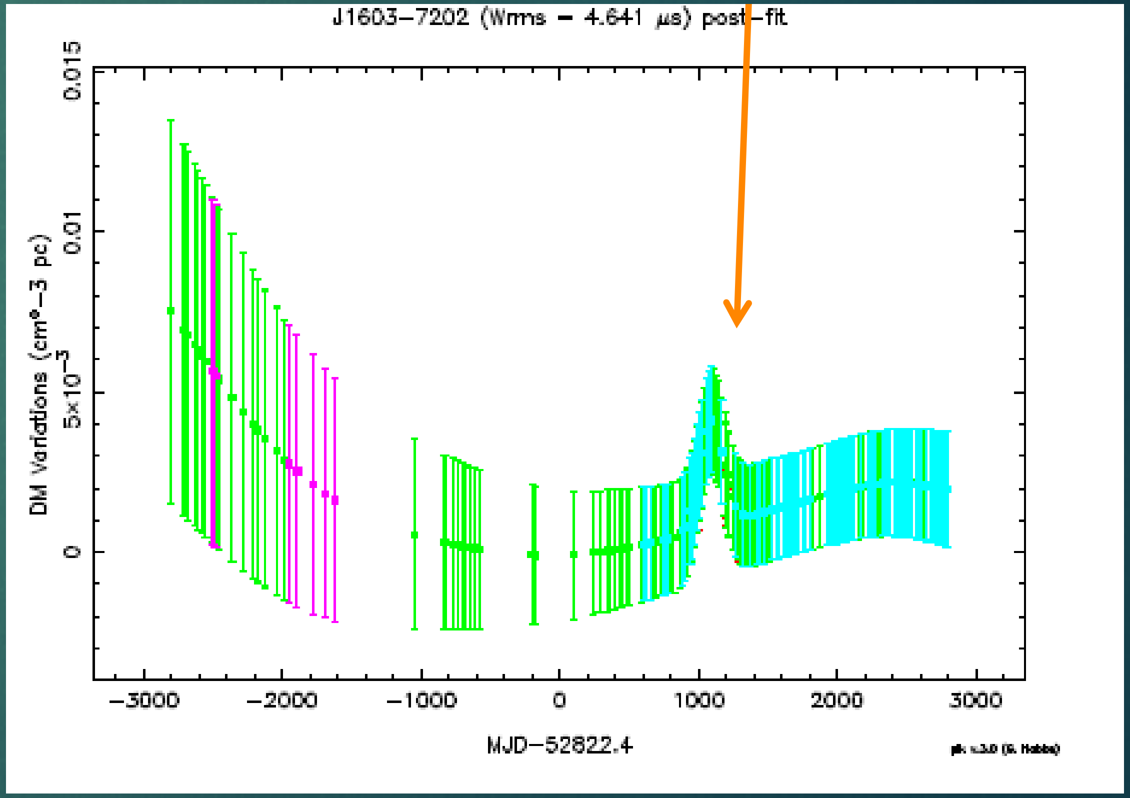
Data Challenges

But the signal isn't stationary…



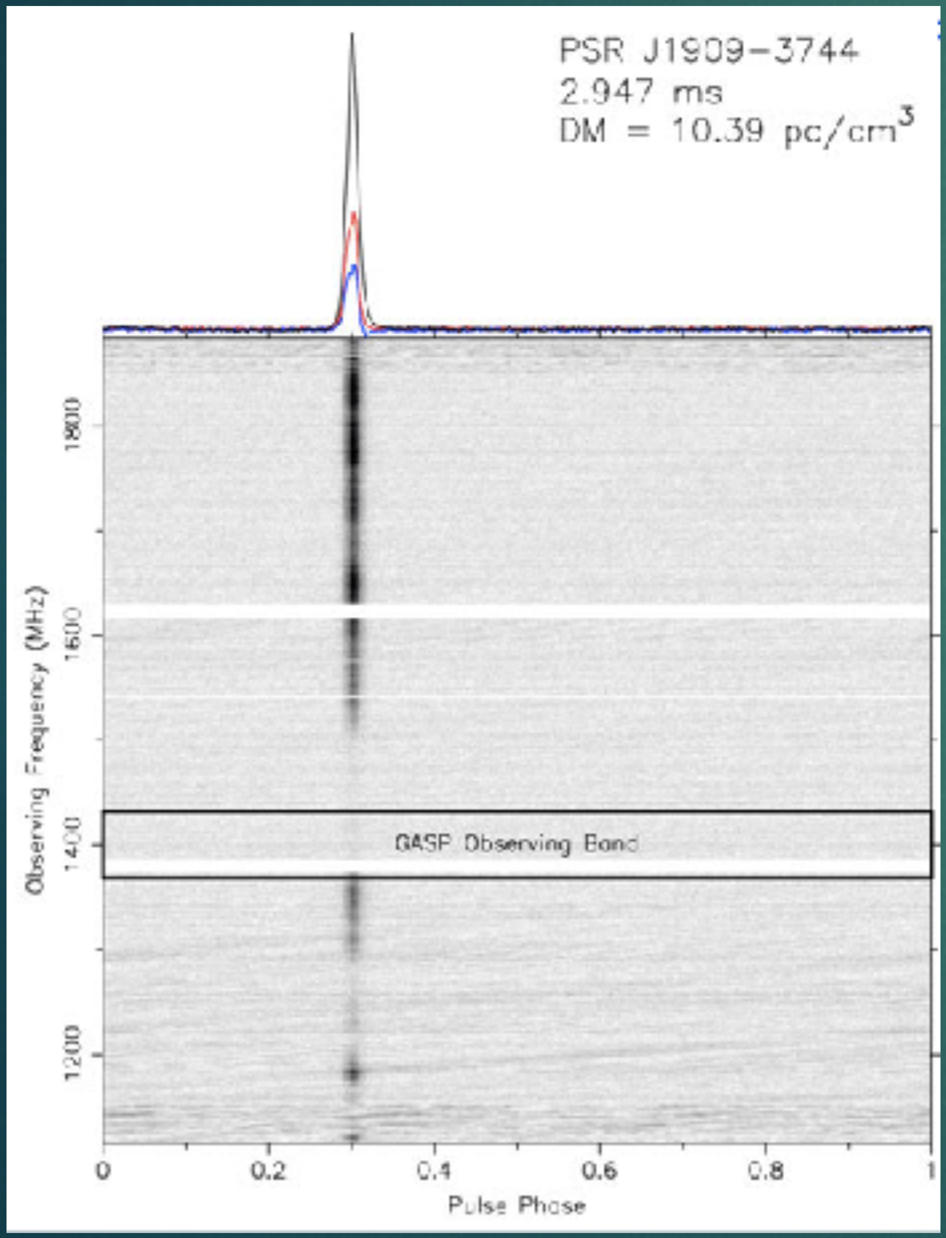
Void in the ISM

Over density in the ISM



Figs: Lentati et al 2016

Data Challenges



So just increase the bandwidth right?

Massive increase over the last few years
Further increases to come

~4GHz simultaneous bandwidth for
up coming systems.

Data Challenges

More than just DM though:
Scattering, 'frequency-dependent DM'

Can really hurt:

PPTA Limits for PSR J1909-3744:
10cm only : 1e-15
10+20cm: 9e-16
10+20+50: 2e-15

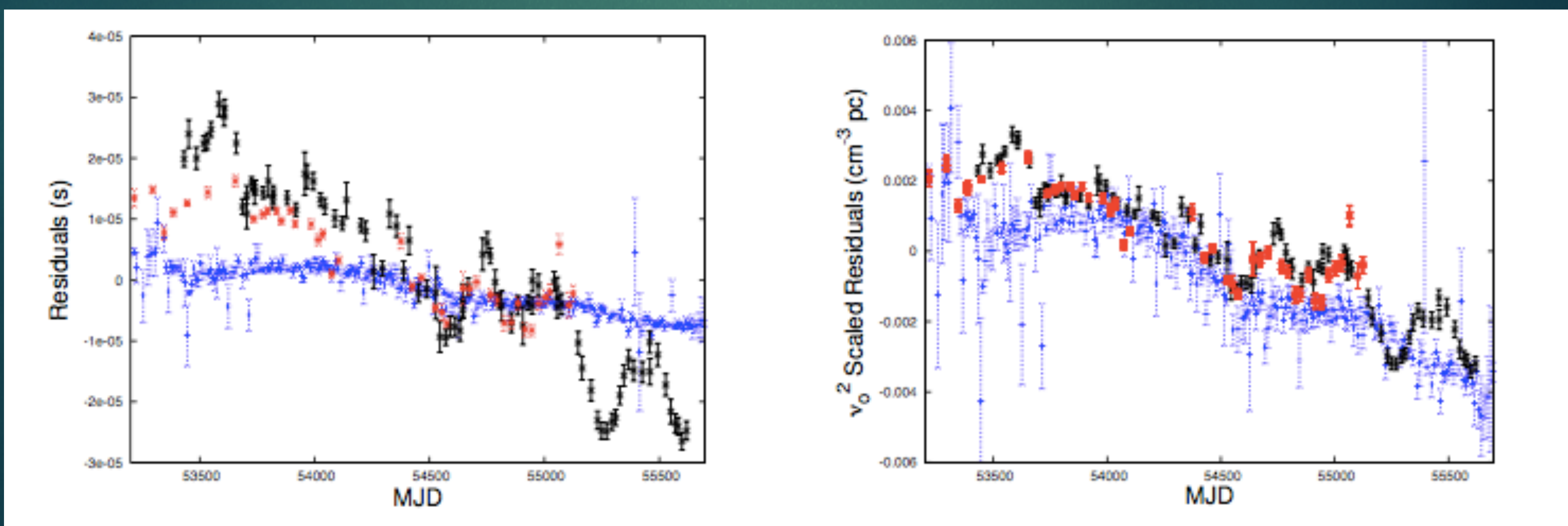


Fig: Lentati et al 2016

Data Challenges

Better modelling can make a huge difference (Lentati et al 2016)
60% increase in sensitivity compared to 'standard' models

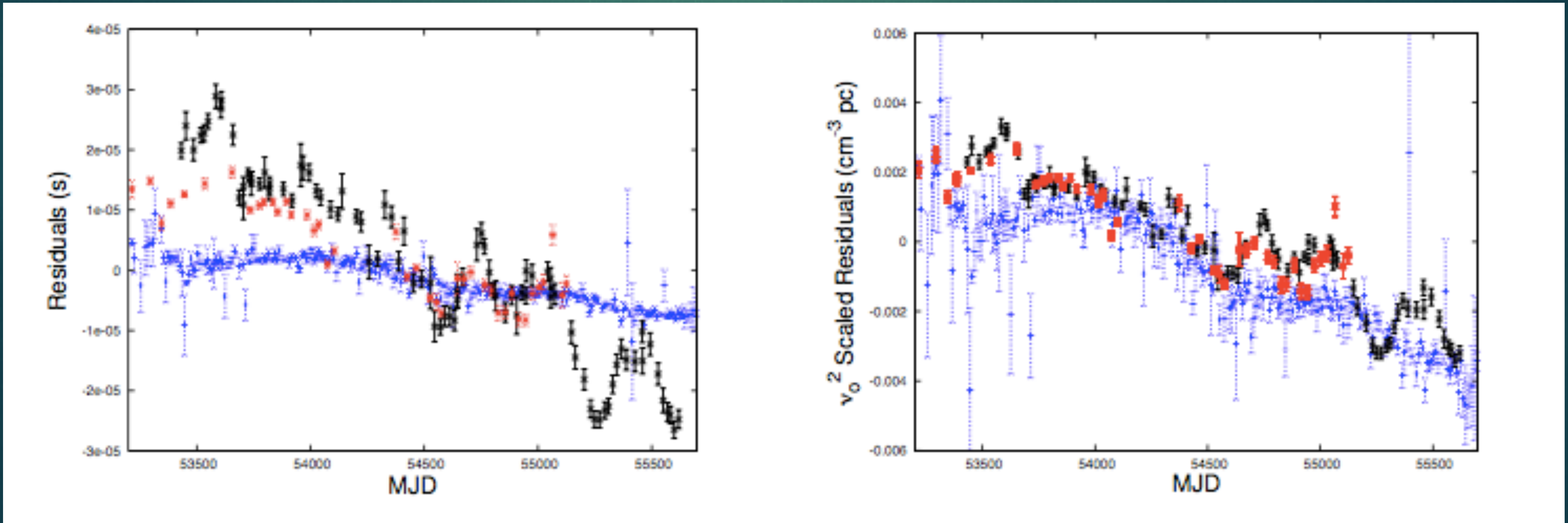
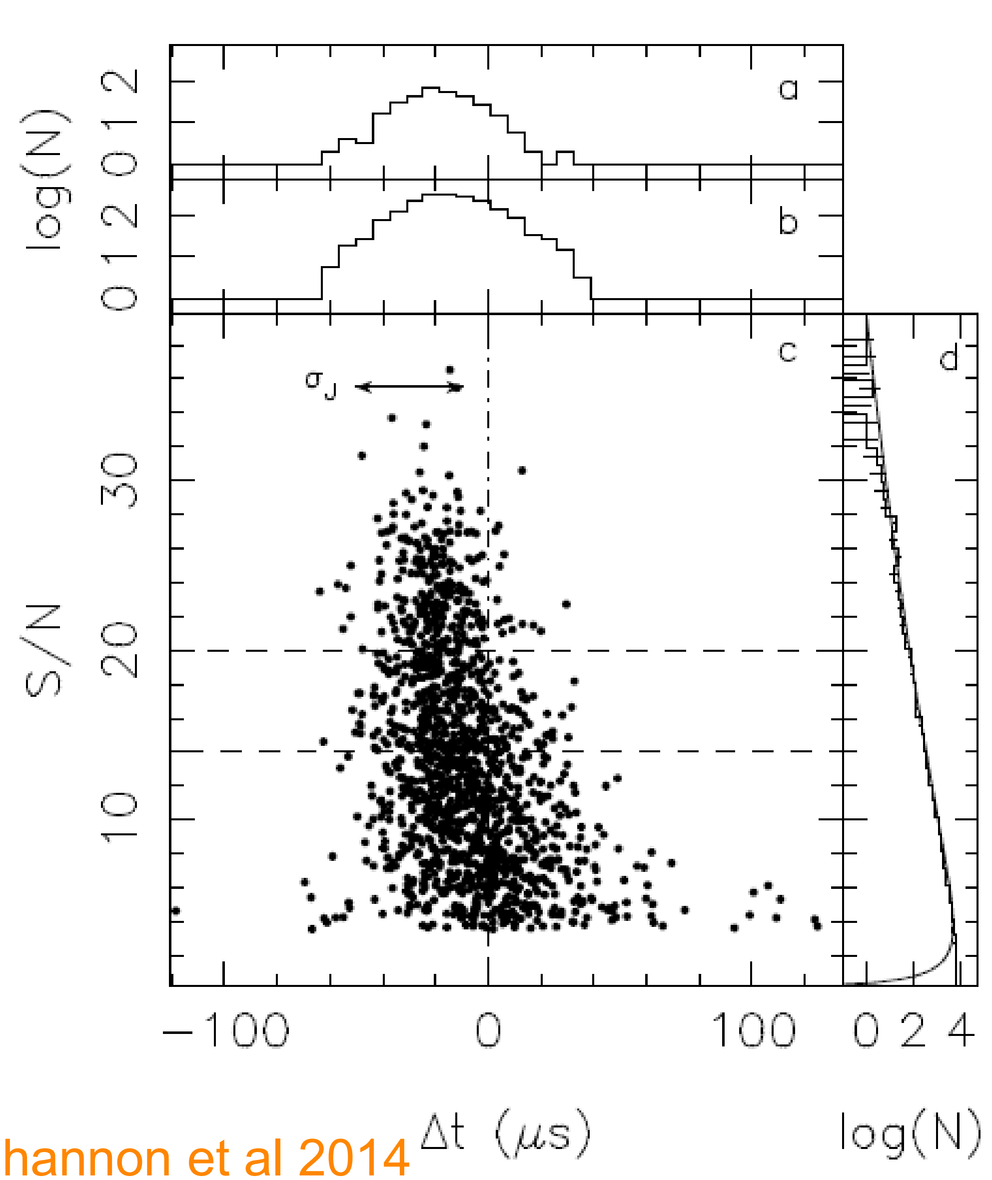


Fig: Lentati et al 2016

Data Challenges



Shannon et al 2014

Intrinsic high frequency variation in arrival time of pulses

Better telescopes won't help.

Already at the limit for some pulsars.

Data Challenges

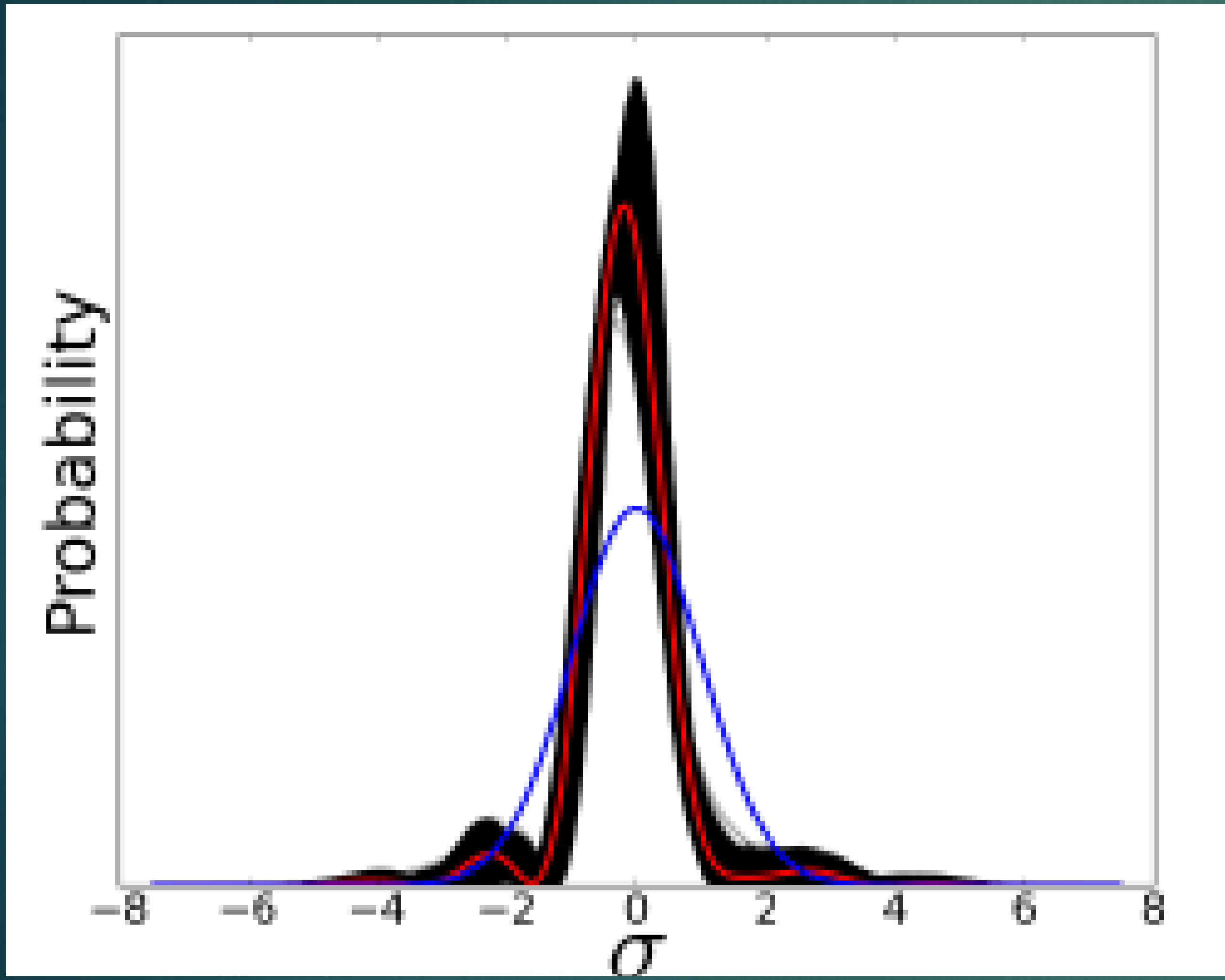


Fig: Lentati et al 2015

Intrinsic high frequency variation in arrival time of pulses

Better telescopes won't help.

Already at the limit for some pulsars.

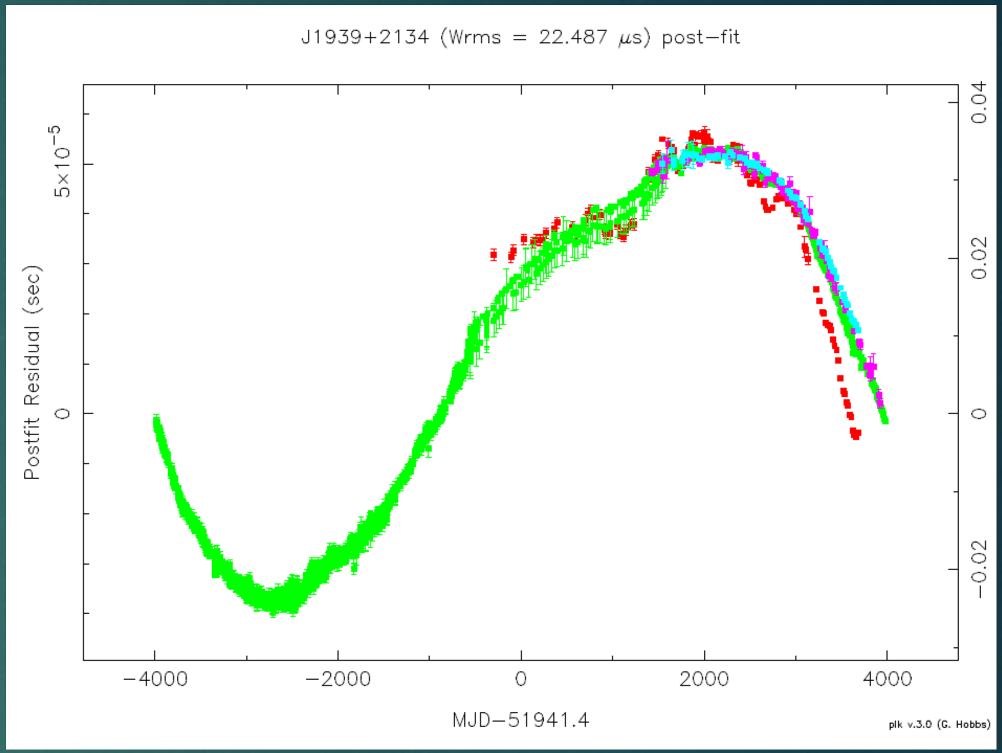
Not necessarily Gaussian either.

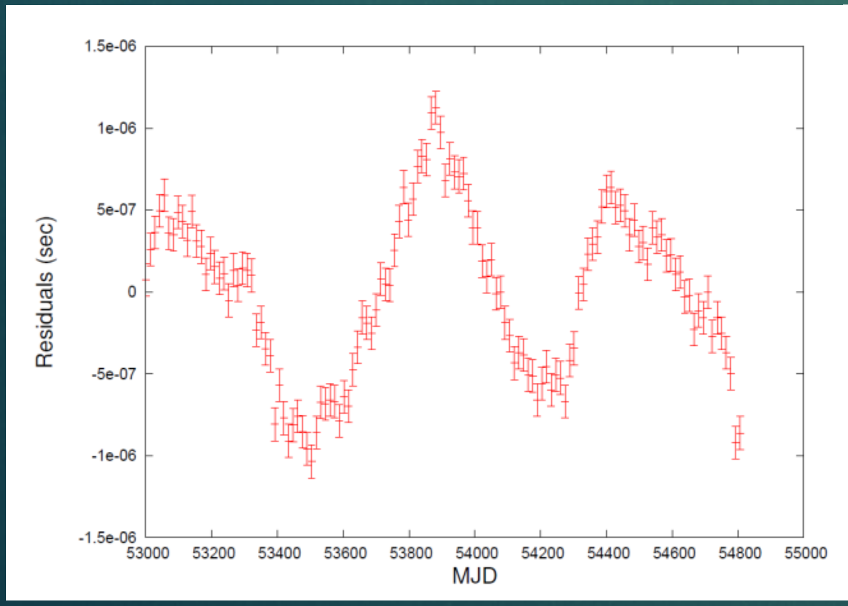
Data Challenges

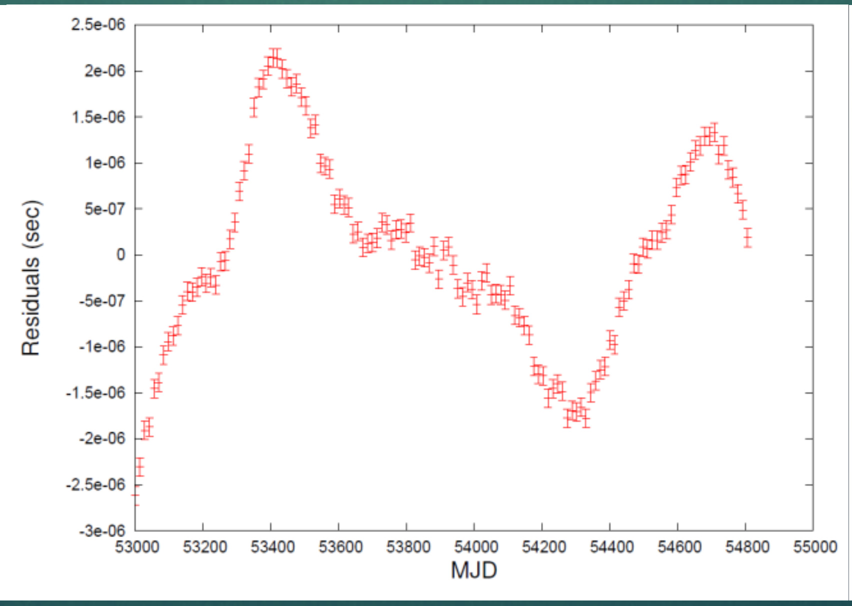
Intrinsic low frequency variation in the arrival times (like Crab) - known as Timing Noise

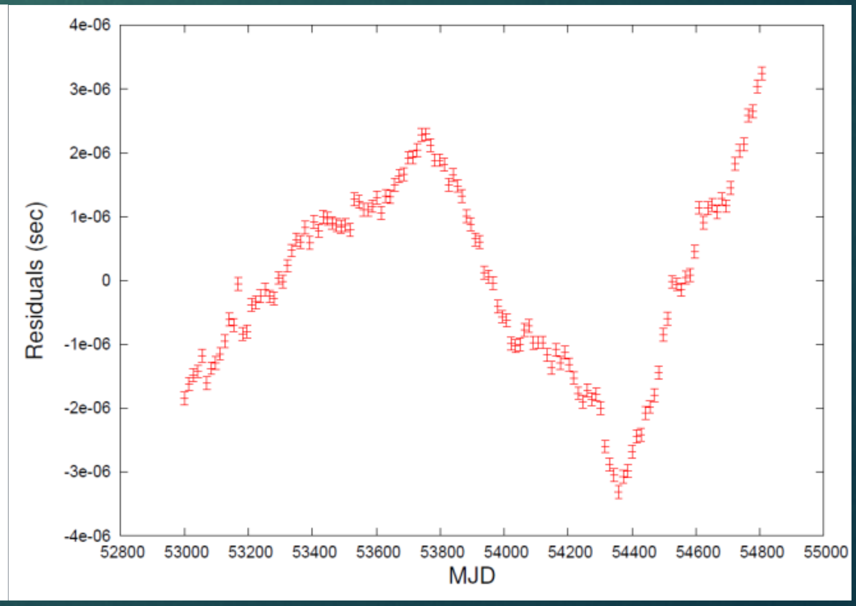
Either from magnetosphere or core… Origins not understood very well.

Stochastic process as with DM - but in one pulsar it can look just like gravitational waves (below).









Data Challenges

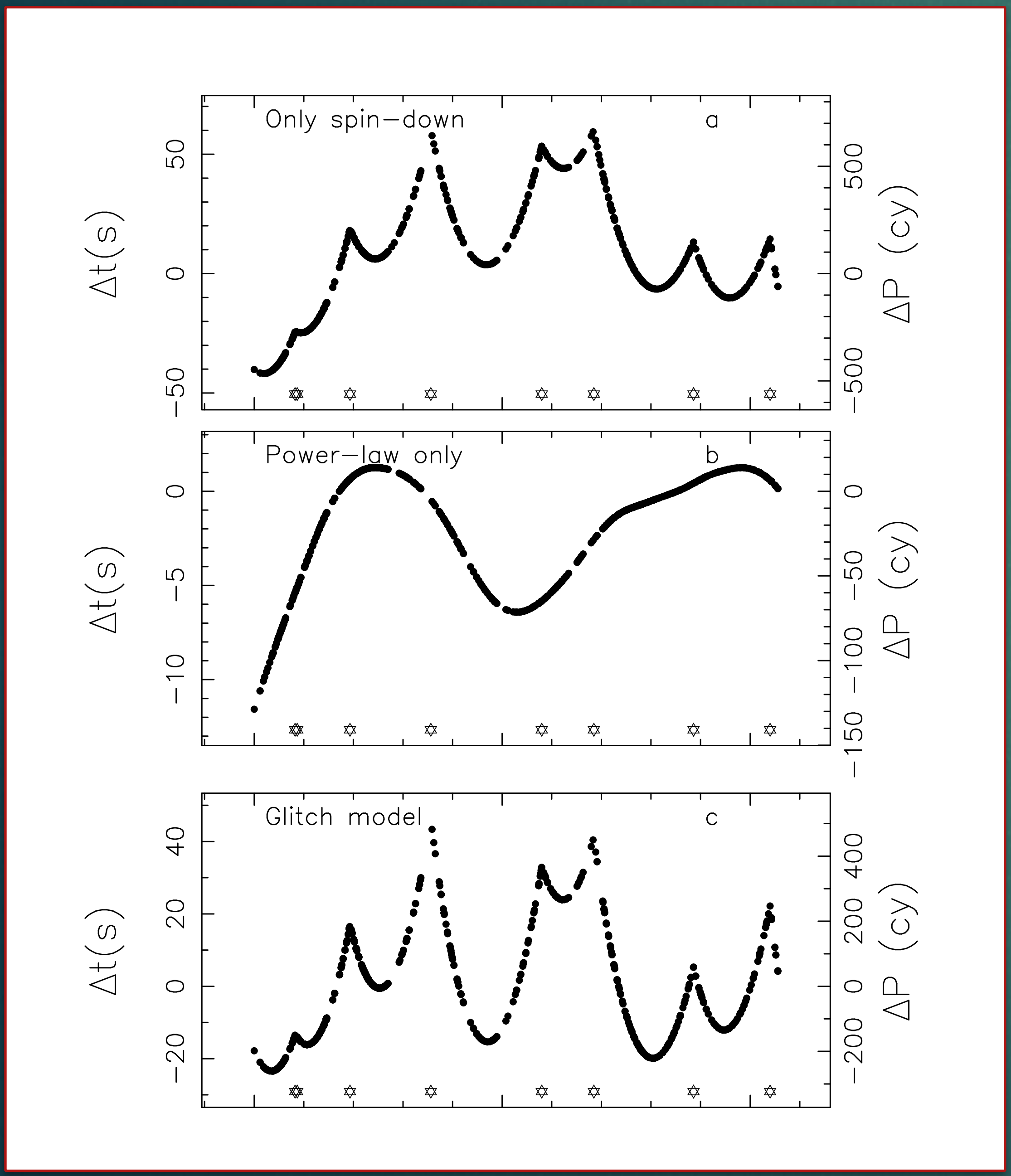


Fig: Shannon et al 2016

Timing Noise from the core:

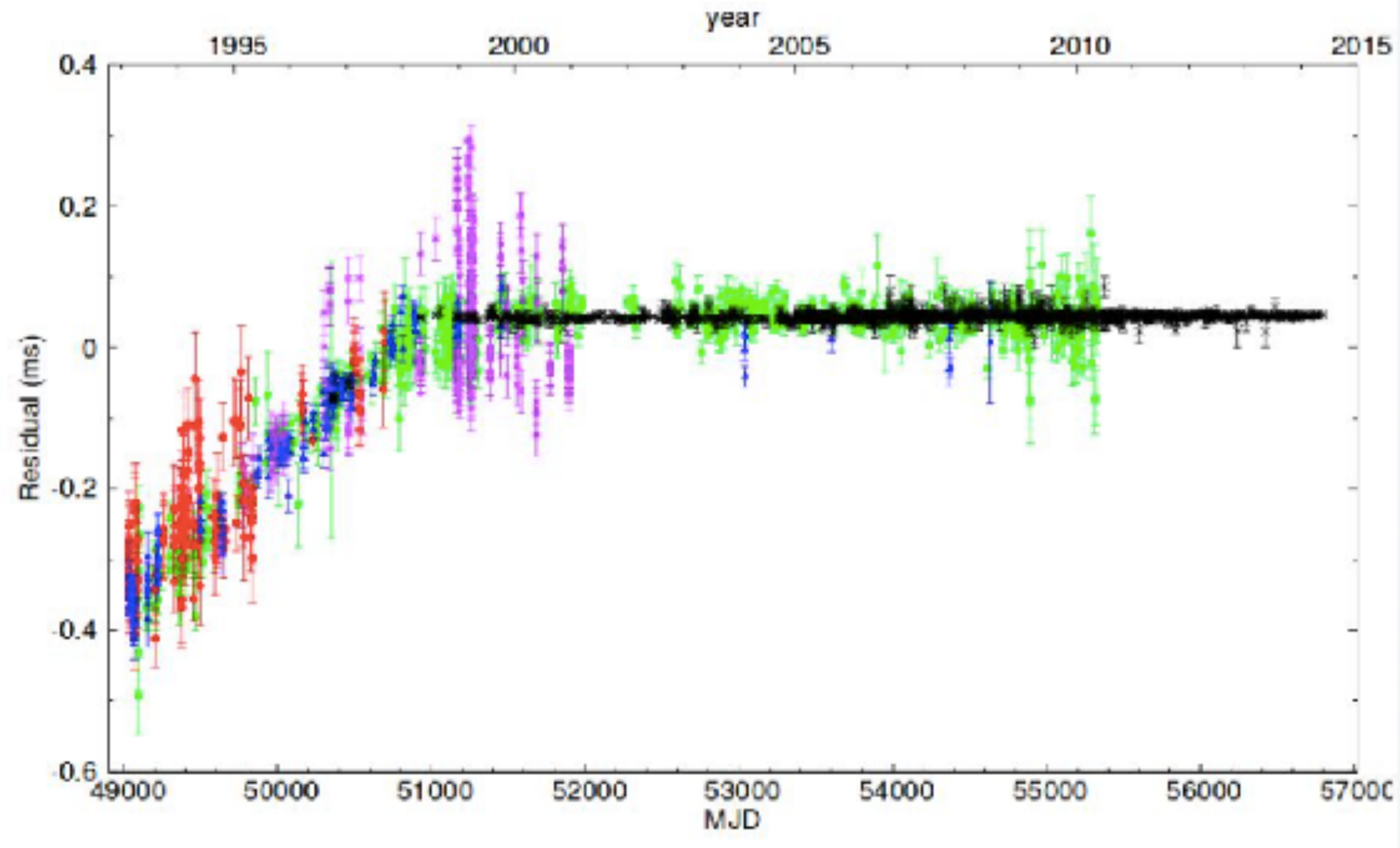
<- Vela (Young slow pulsar)

Glitches - sudden changes in rotation rate
Accompanied (in this case)by long
(~1000 day) decays

Maybe associated with the transfer of angular momentum between the superfluid interior and solid crust of the neutron star.

Common in young pulsars
But two glitches found in millisecond pulsars

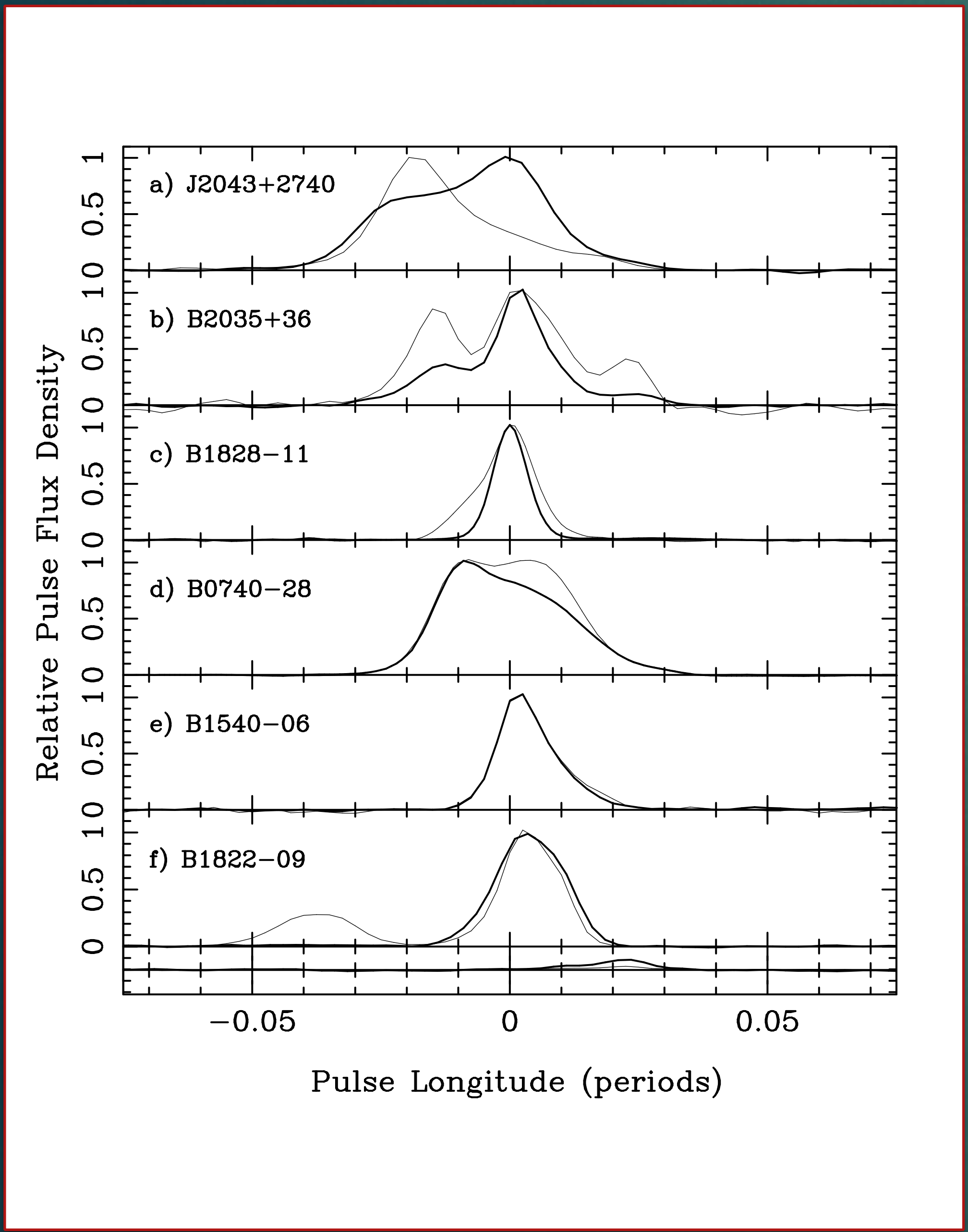
Data Challenges



Glitch in the MSP
J0613
McKee et al 2016

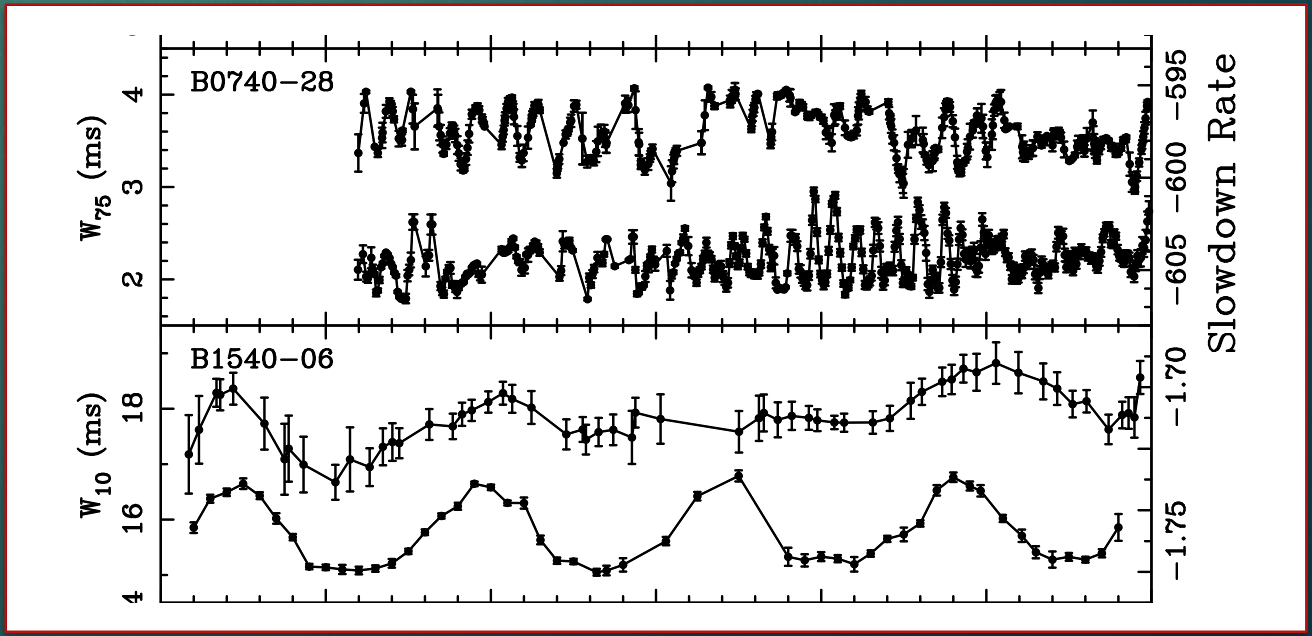
Sounds like bad news?
Glitches are not so hard.
Put it in the model, decreases long term sensitivity,
but at least somewhat deterministic.

Data Challenges



Timing Noise from the magnetosphere:
Less extreme: Switching to different states

Observe change in pulse shape:
Rate of energy loss is different
different spin down rate



Figs: Lyne et al 2010

Data Challenges

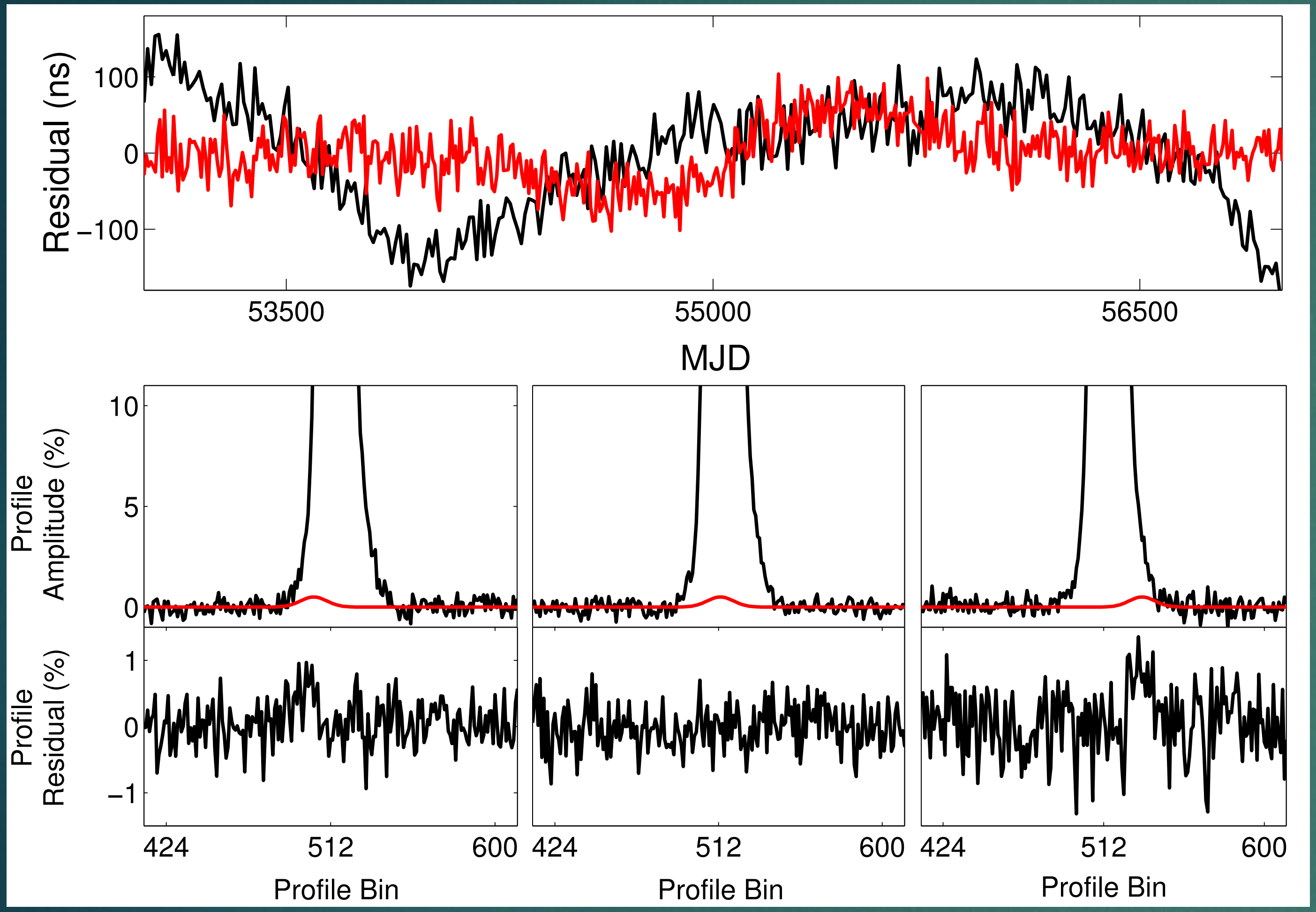


Fig: Lentati & Shannon 2015

But:
Profile change can lead to 'timing noise' in the arrival times due to mismatch between template and profile data.

<- Simulation

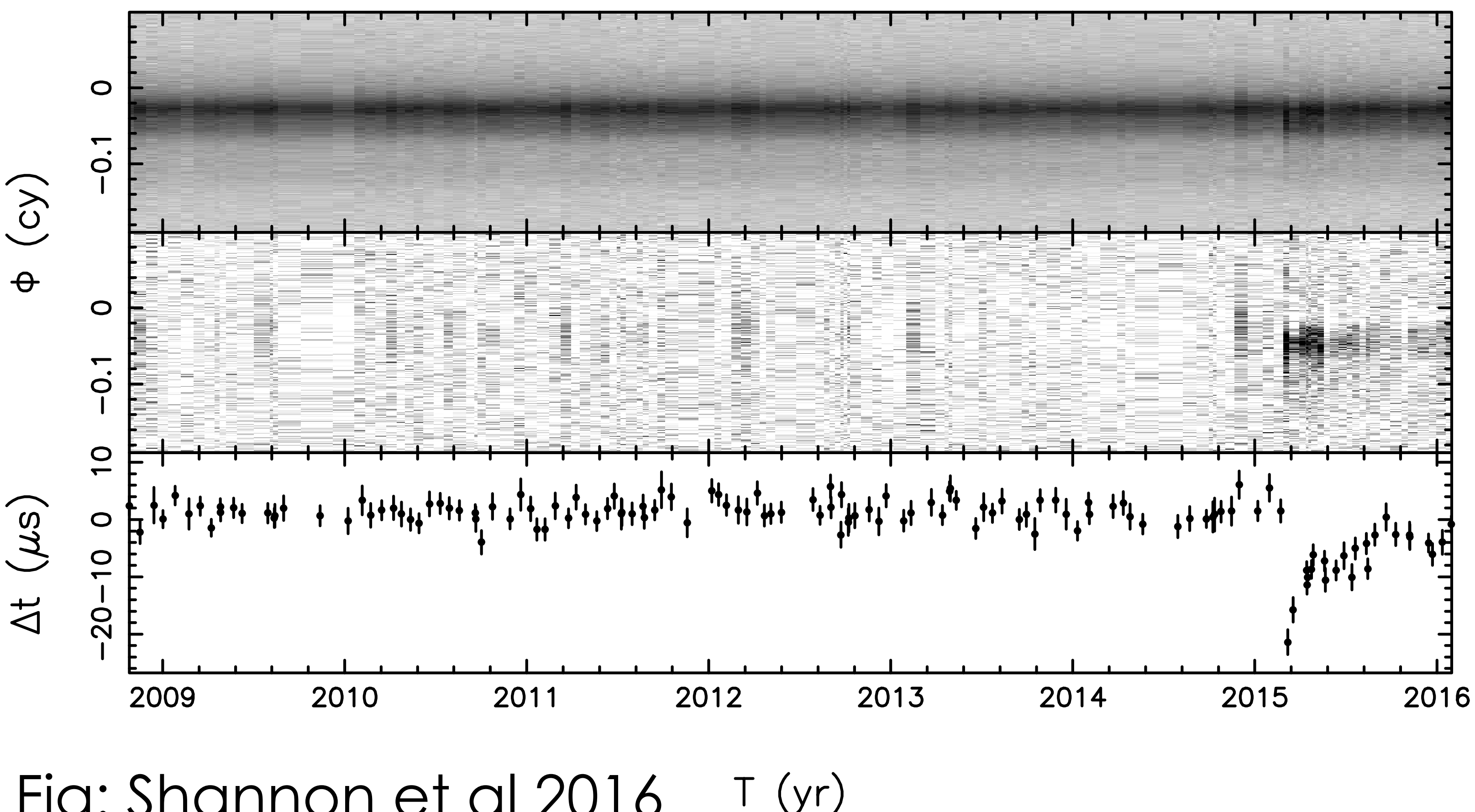
Change in pulse shape lead to observed timing noise when comparing profile to stationary model.

Black curve = signal from GWs at current upper limit.

Red = residual induces from < 1% change in profile shape

Data Challenges

Time-correlated profile change seen in young pulsars a lot
Recently seen in a millisecond pulsar too.
The shift in the residuals isn't an actual shift. Just mismatch between template and data. (Shannon et al 2016, Liu et al 2015)



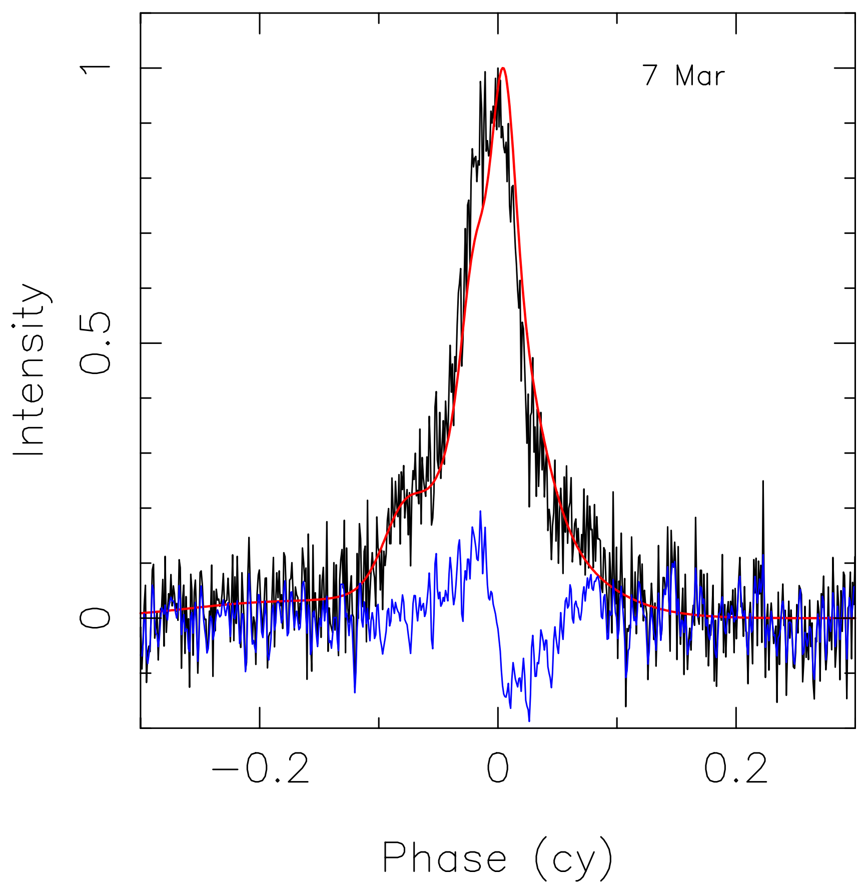


Fig: Shannon et al 2016

Data Challenges

Different approach: Profile domain timing
Don't make time of arrivals.
Simultaneously estimate model for profile and pulsar timing parameters.
Decouple shape change from shifts.

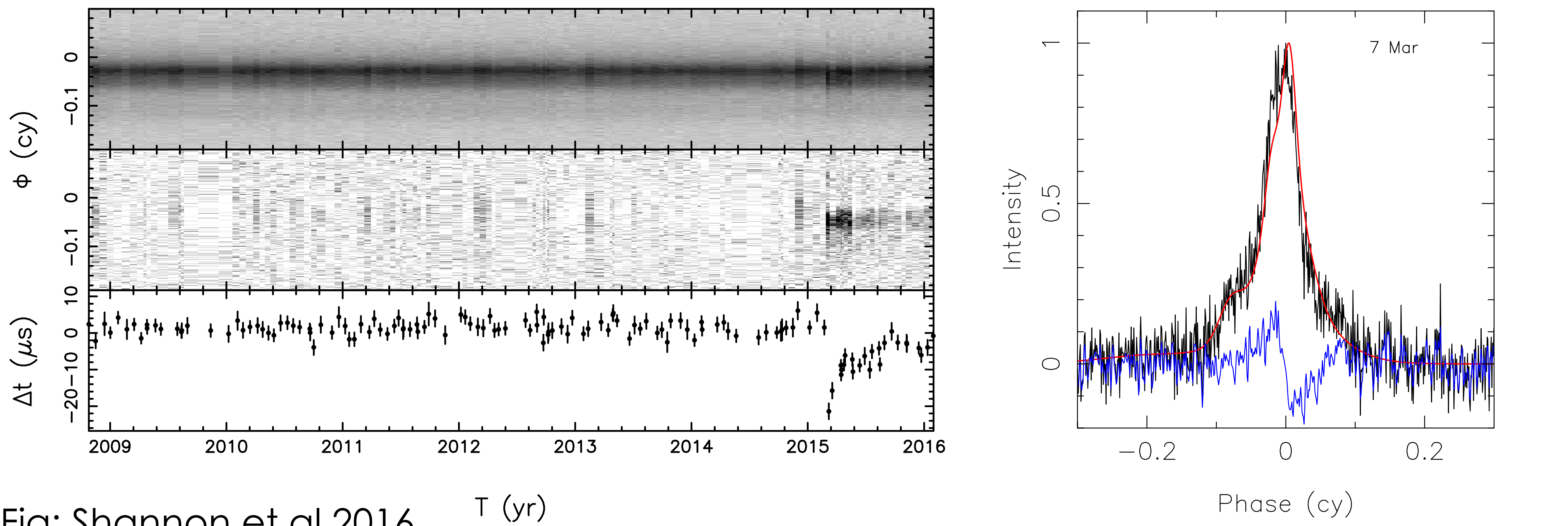
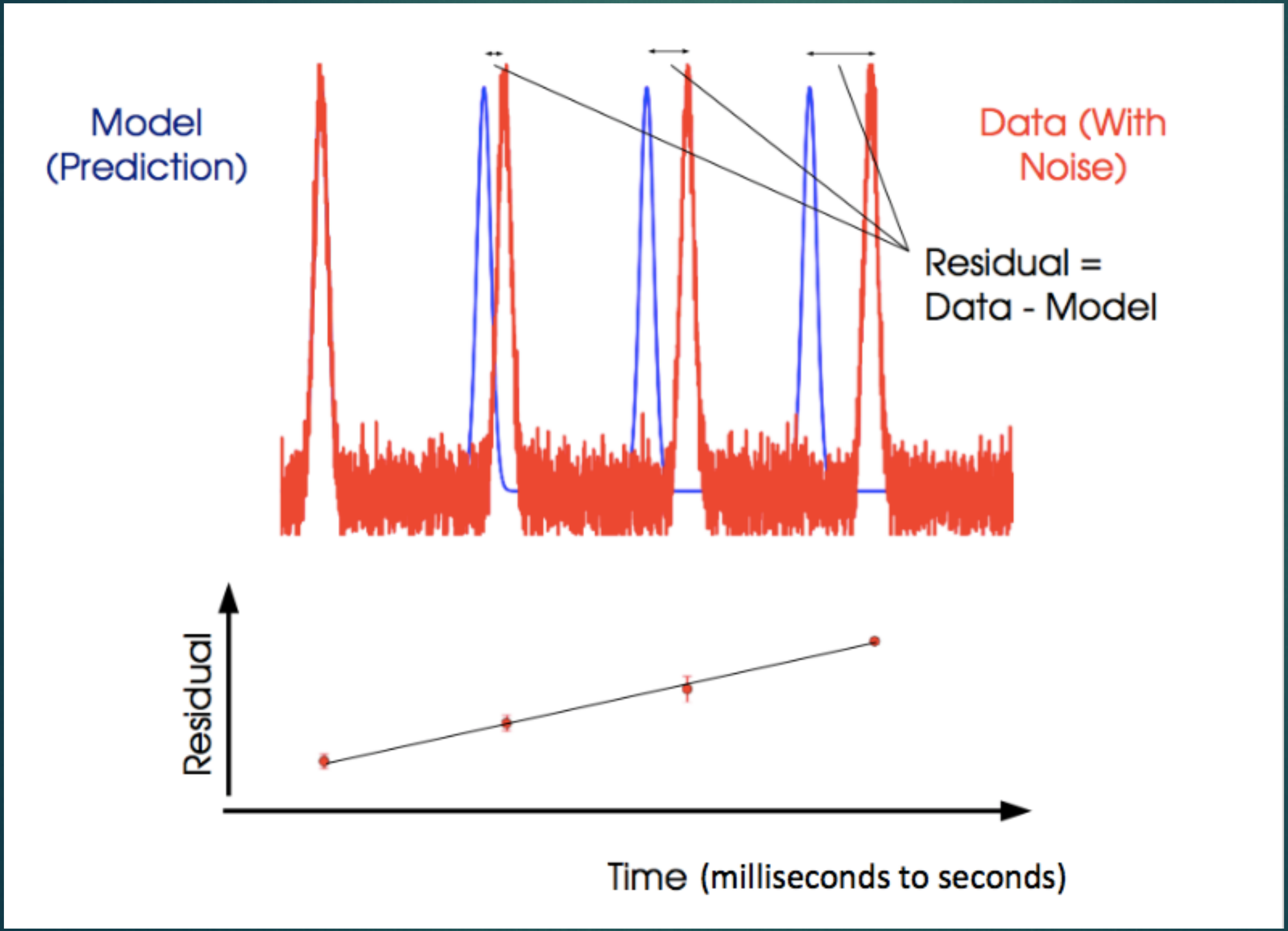
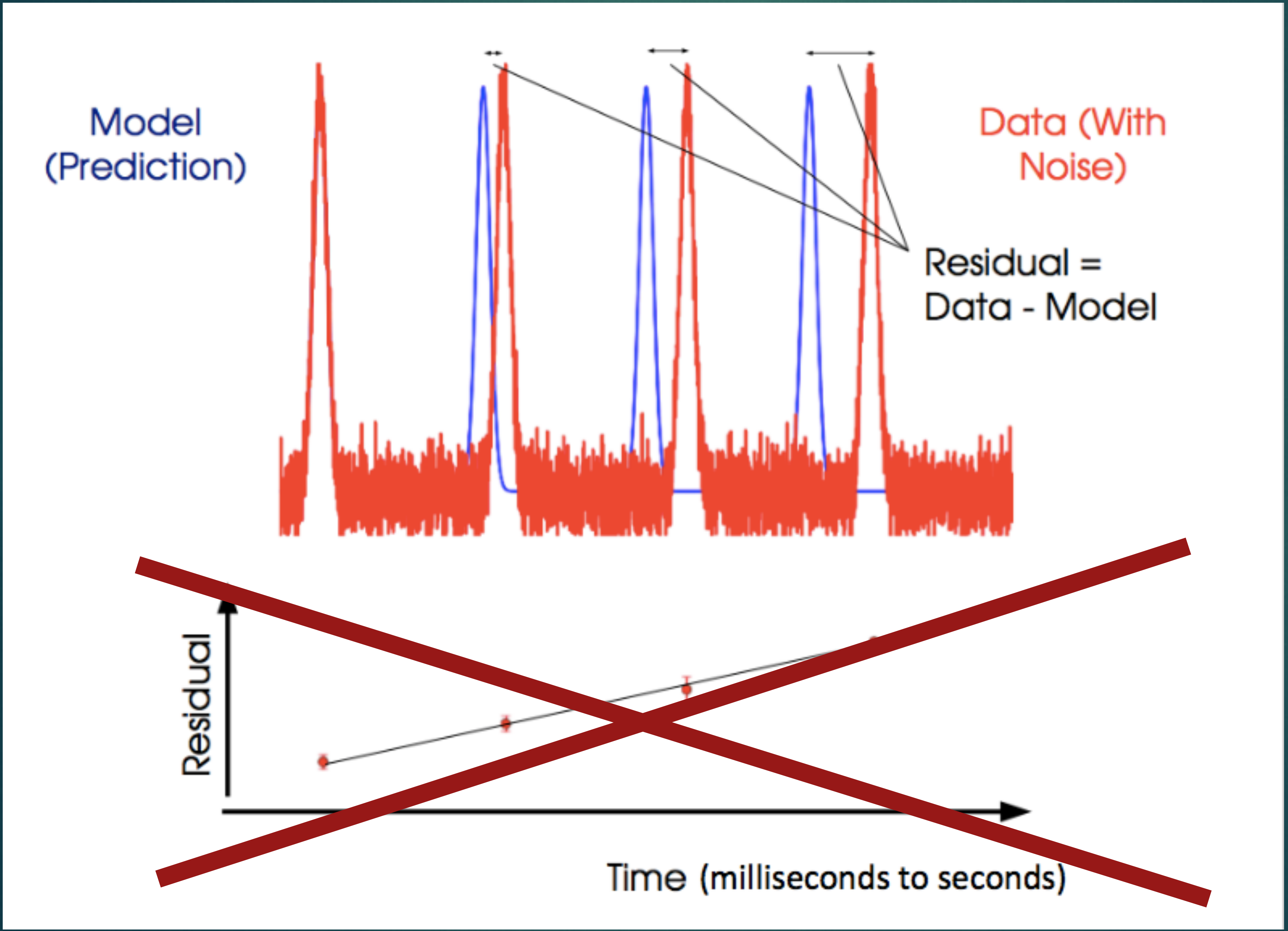


Fig: Shannon et al 2016

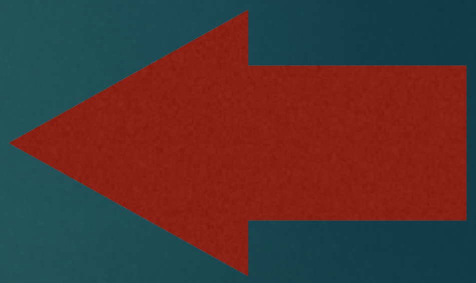
From Earlier This Week:



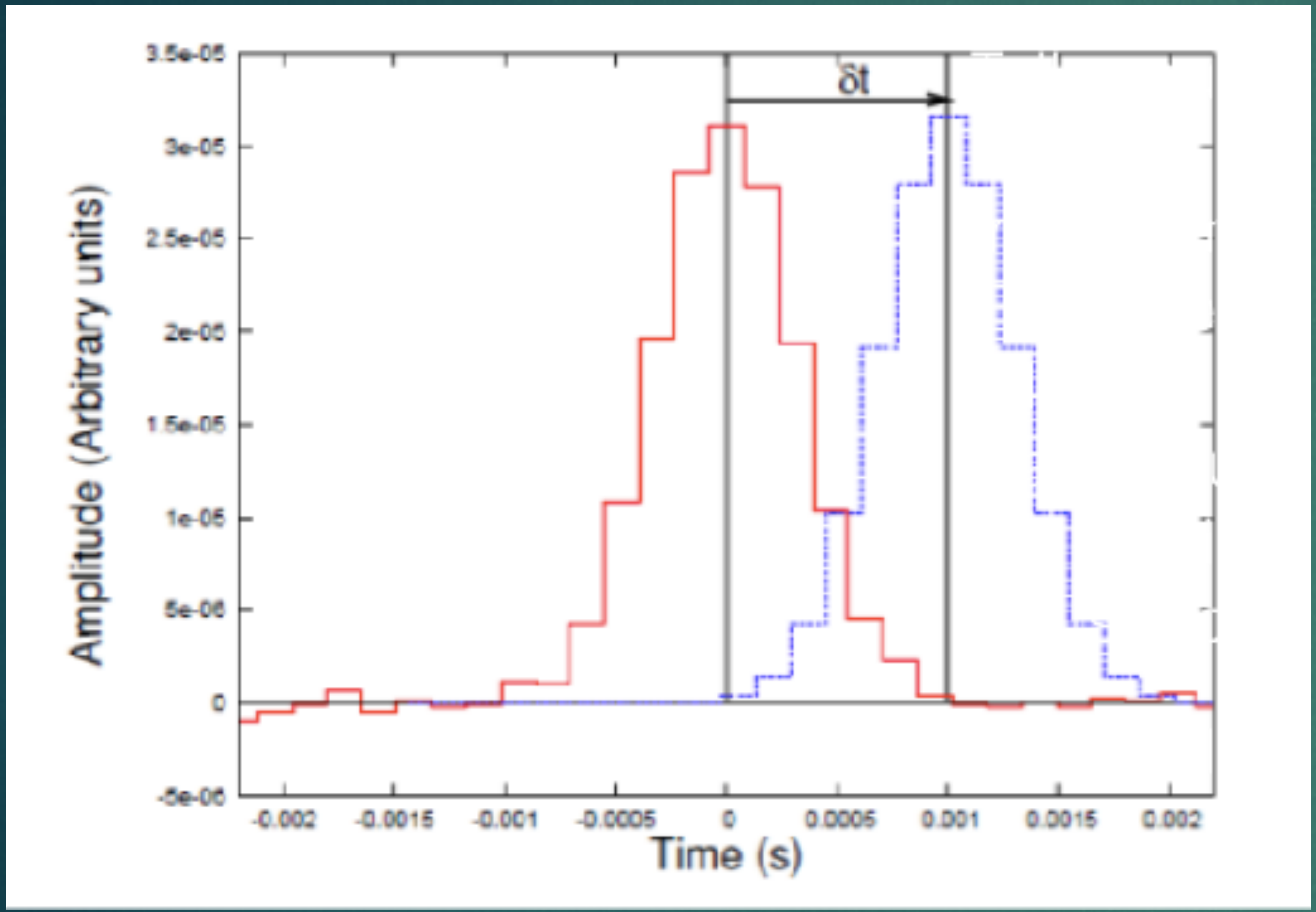
Don't have to make ToAs.
Just work directly with the profile data.



Stop Here



Profile Domain Timing



Sample from the Timing Model:
Tells you when you expect your pulses

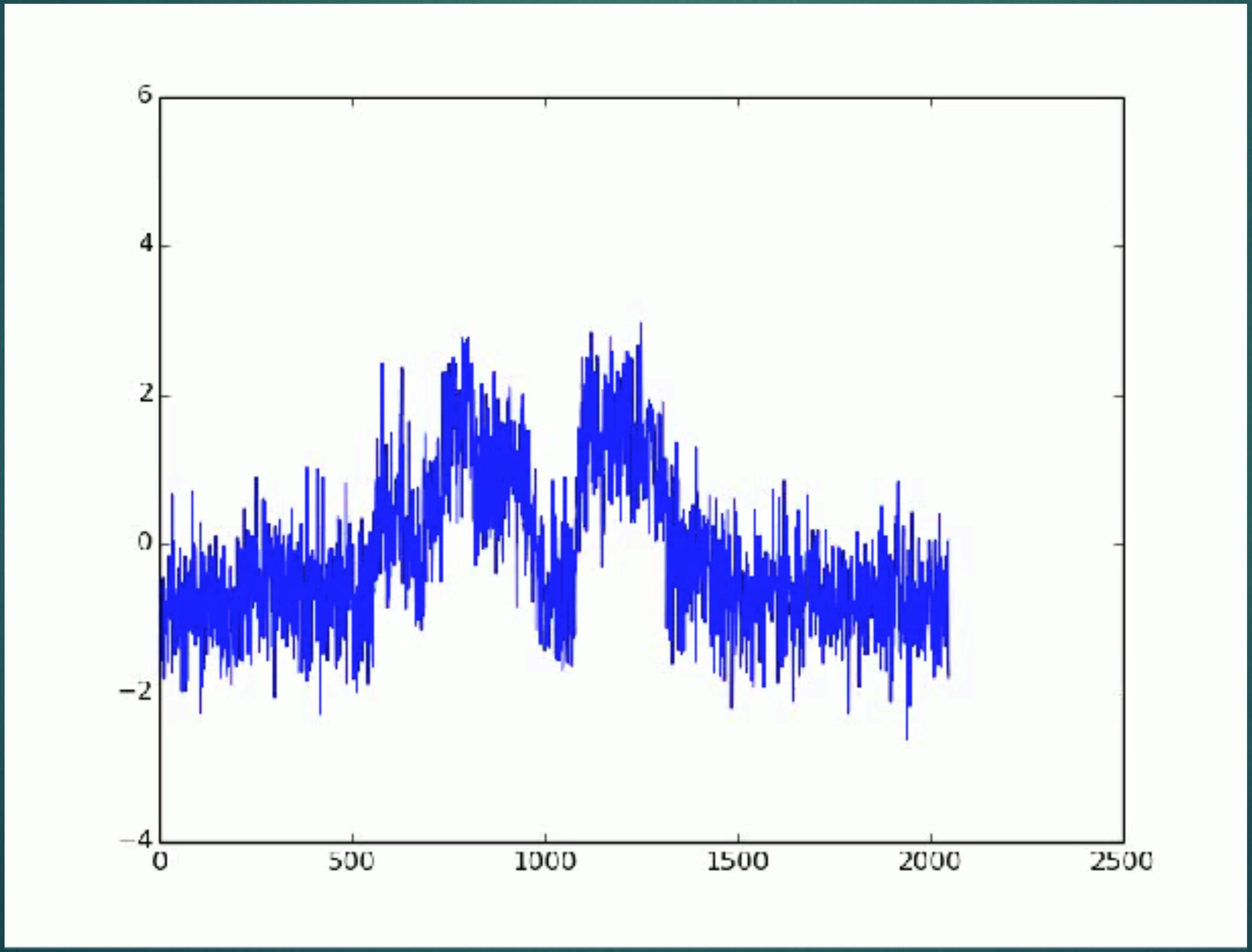
Include model for profile:
Evaluate this where your timing model tells you to.
Can include:
profile evolution,
shape variation

Log-Likelihood is then:

Sum over Profiles : (Profile data - Profile model)^2/(Profile noise)^2

E.g. PSR J0437

Profile Residuals: Very significant in most observations.



Data Challenges

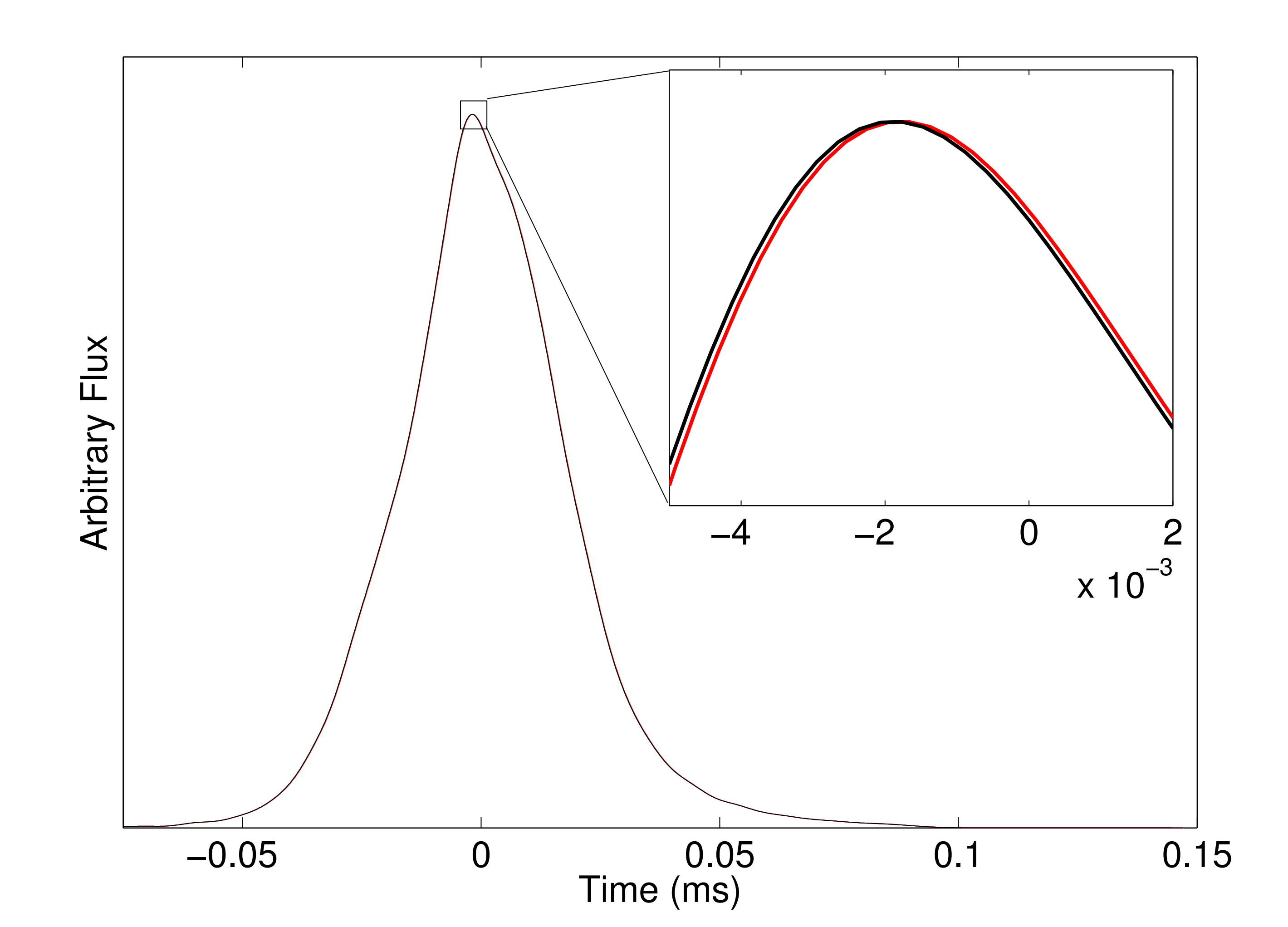
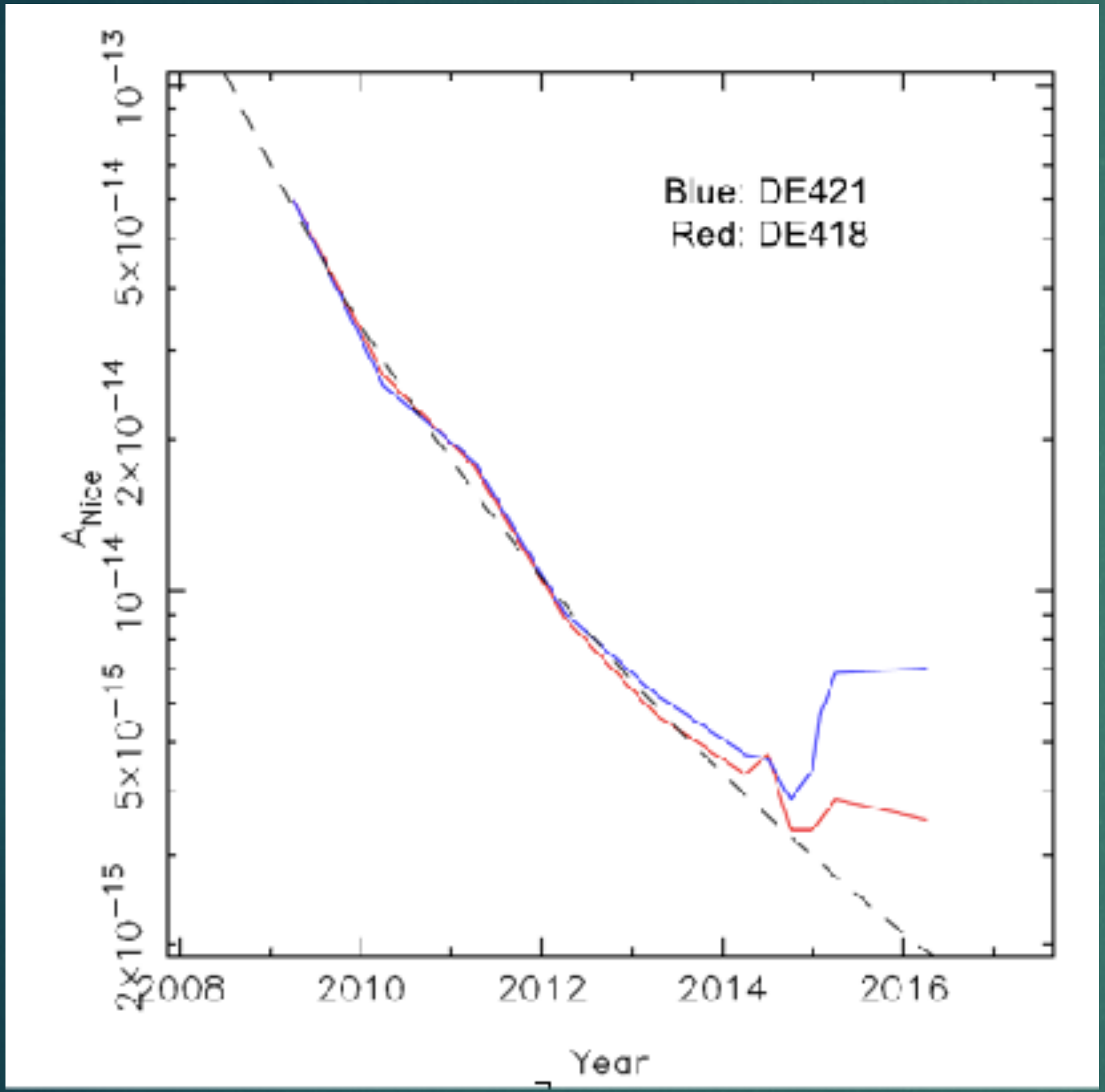


Fig: Lentati & Shannon 2015

Need to be accurate: Shift due to GWs is only a tenth of a phase bin.

Standard timing approach makes it difficult/impossible to distinguish timing noise due to shifts, from timing noise due to changing profile and mismatched template.

Data Challenges (Last One)



PPTA limit as a function of time:
Dashed line = Theoretical decrease for noise only
Different colours are different models for the Solar System (JPL Ephemeris)

Limits now depend on this :(

Fig: Ryan Shannon

Data Challenges (Last One)

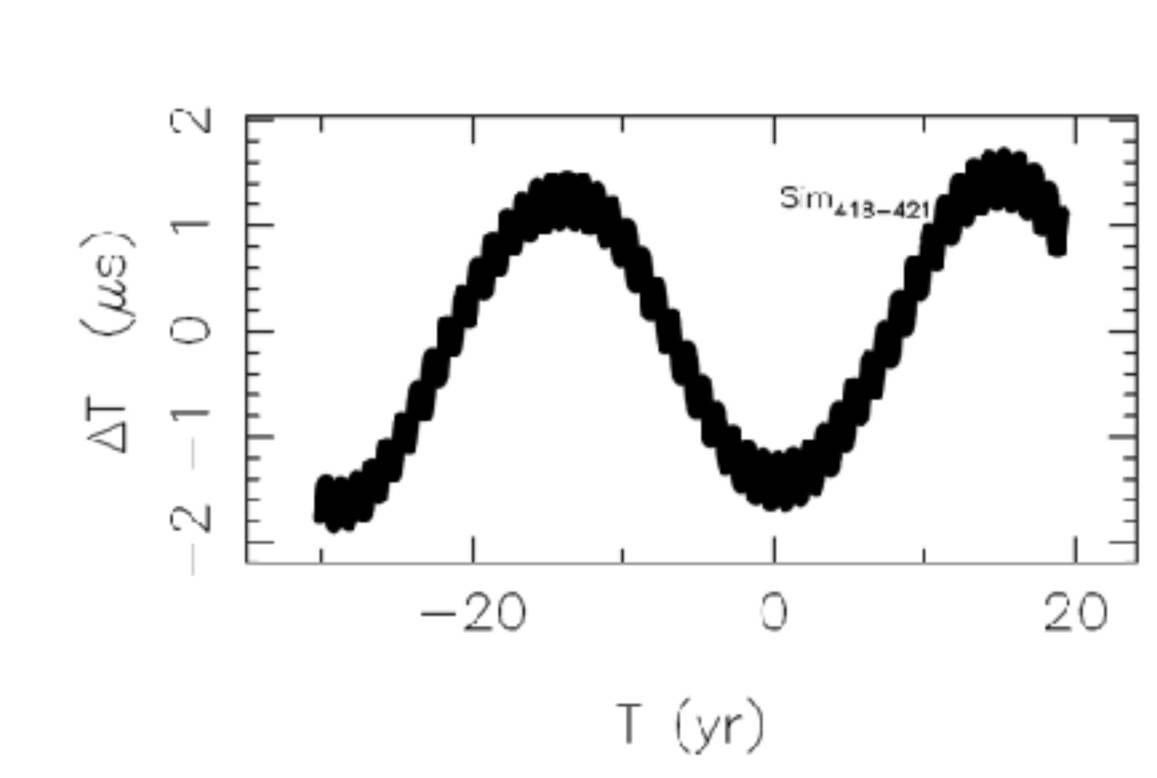


Fig: Ryan Shannon

Simulated arrival times over > 40 years

Simulated in DE418 and measured in DE421

Looks like Saturn..

Cool! But Annoying..

One (Unrelated) Last Thing!

Disneyland on Saturday?

On to the workshop!