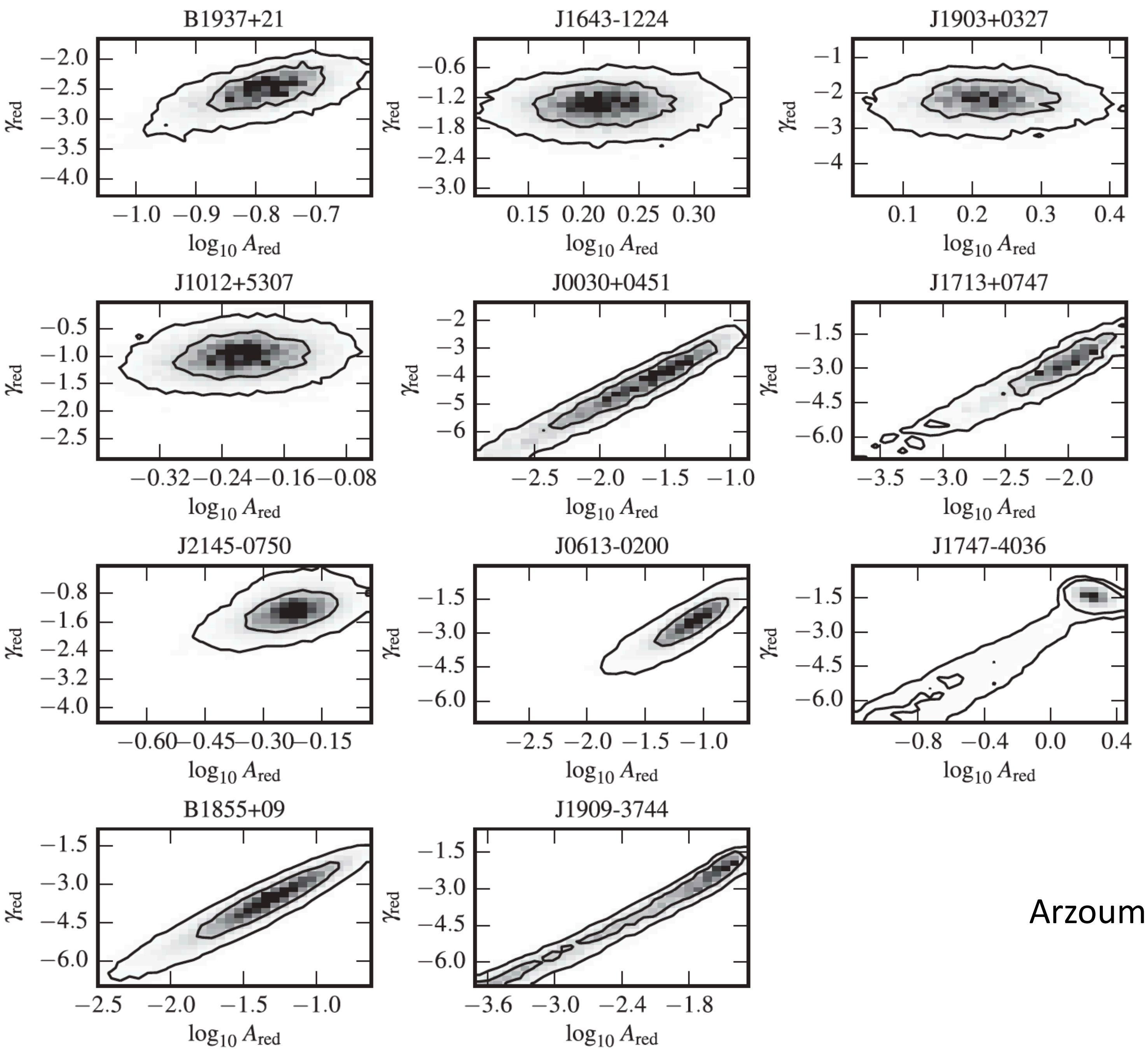
Noise and noise modelling in pulsar-timing observations

Ryan Shannon (Swinburne/Ozgrav)



Outline

- Physical model
- Signal model
- Bayesian methods

Arzoumanian et al. (2018)

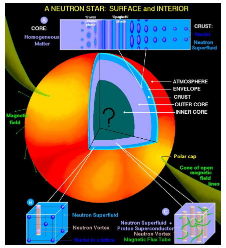
Figure 2. 2D posteriors of the amplitude and spectral index of the red-noise parameters for those pulsars with Bayes factors for red noise greater than 100, plus J1713 +0747. The contours are the 50% and 90% credible regions.

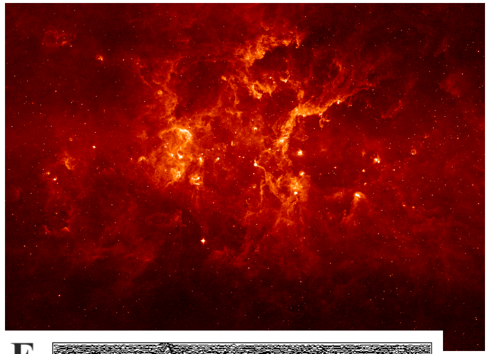
Contributions to Pulsar Arrival Times

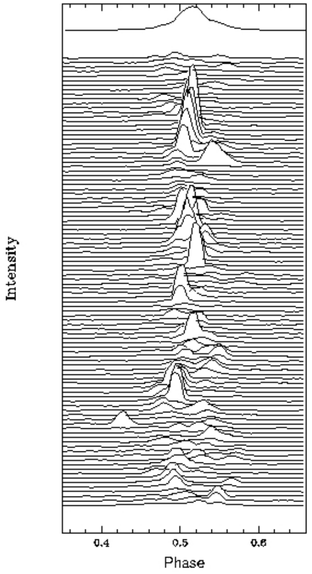
Pulsar

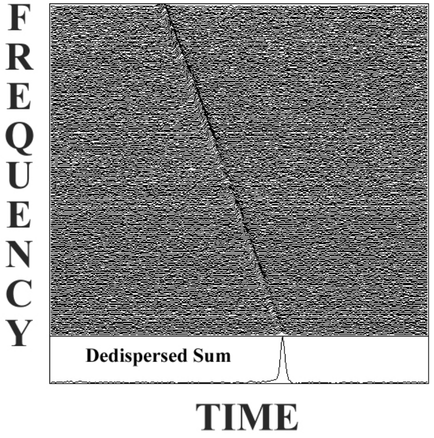
- Light travels 30cm in 1 ns
- Pulsar spindown
- Intrinsic variation in shape and/or phase of emitted pulse
- ***Reflex motion from companions.***
- Pulsar position, proper motion, distance
- ***Stochastic spindown variations***
- ***Gravitational Waves***
- ***Warm electrons in the ISM***
- Solar system

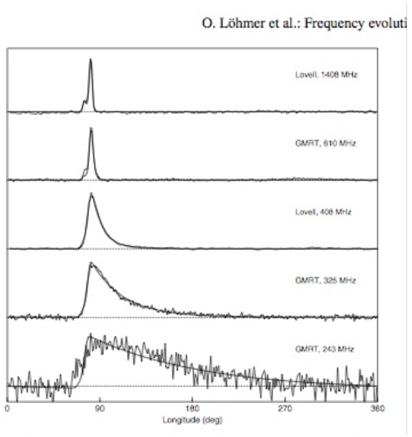
Earth





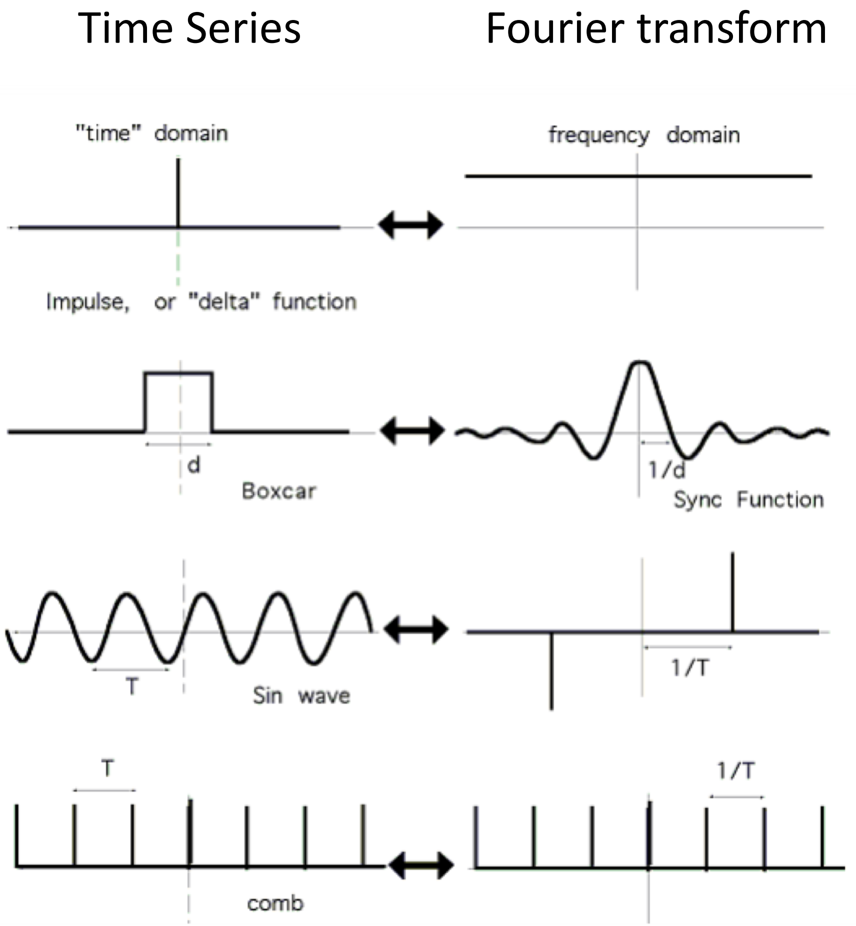






Power spectral estimation in one slide

- Oftentimes signals are easier expressed in the "frequency domain" rather than the time domain
  - Sine wave can be expressed by three numbers:
    - Amplitude
    - Frequency
    - Phase
- Fourier transform:
  - Projection of time series onto set of sinusoids
  - Fast Fourier transform (FFT): efficient algorithm of calculating FT
- Power spectrum
  - Square of Fourier transform
  - Useful if don't care about phase of signals
- We'll be using frequencies in three different contexts
  - Frequency = fluctuation frequency
  - Frequency = radio frequency
  - Frequency = gravitational wave frequency



Deterministic vs. Stochastic Process

- **Deterministic system:**
  - If initial state is known exactly, future state can be predicted
  - E.g. Pulsar, white dwarf system

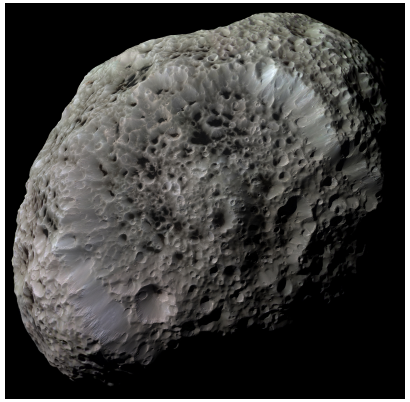
- **Chaotic system (type of deterministic system)**
  - If deterministic system behavior is highly sensitive to initial condition, evolution will show apparent randomness
  - Example: Hyperion

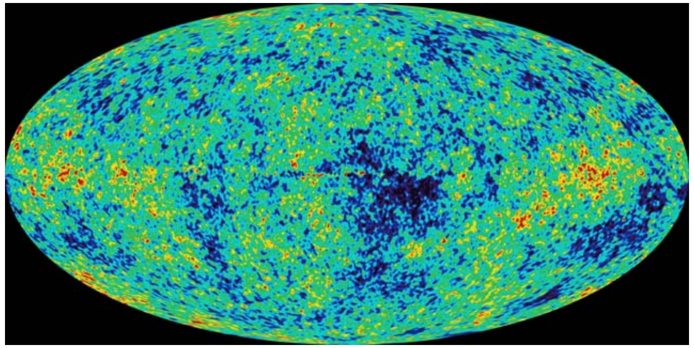
- **Stochastic system:**
  - System evolution depends on random variables
  - Example CMB, Gravitational Wave Background

- A large portion of the rest of the talk will be devoted to identifying, assessing, and characterizing stochastic processes.

Table 1 PSR J0437–4715 physical parameters

Right ascension, α (J2000)	$04^{h}37^{m}15^{s}.7865145(7)$
Declination, δ (J2000)	$-47°15'08''.461584(8)$
μ_α (mas yr^{-1})	121.438(6)
μ_δ (mas yr^{-1})	-71.438(7)
Annual parallax, π (mas)	7.19(14)
Pulse period, P (ms)	5.757451831072007(8)
Reference epoch (MJD)	51194.0
Period derivative, $\dot{P}$ (10^{-20})	5.72906(5)
Orbital period, P_b (days)	5.741046(3)
x (s)	3.36669157(14)
Orbital eccentricity, e	0.000019186(5)
Epoch of periastron, T_0 (MJD)	51194.6239(8)
Longitude of periastron, ω (°)	1.20(5)
Longitude of ascension, Ω (°)	238(4)
Orbital inclination, i (°)	42.75(9)
Companion mass, m_2 ($M_\odot$)	0.236(17)
$\dot{P}_b (10^{-12})$	3.64(20)
$\dot{\omega}$ (° yr^{-1})	0.016(10)

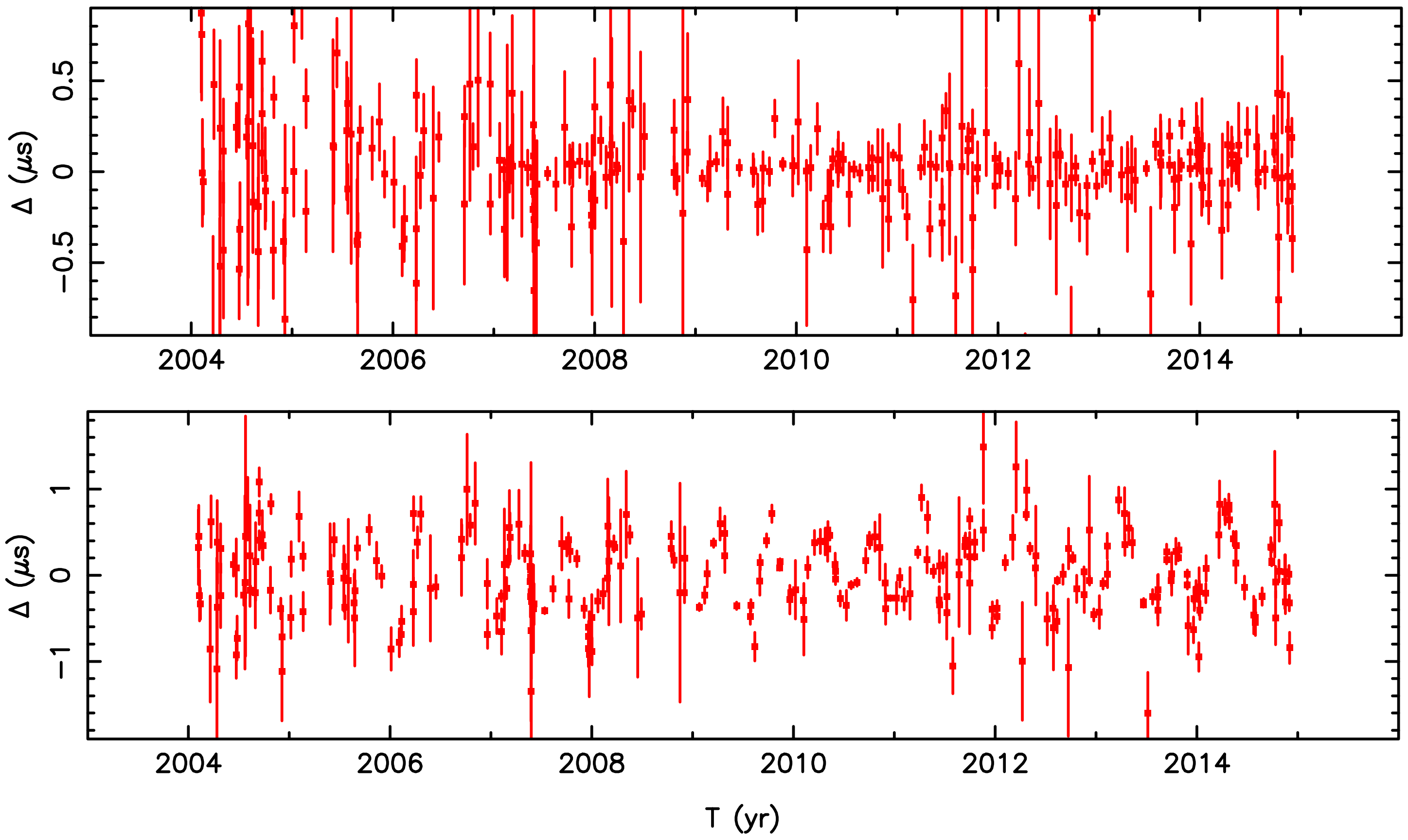




Example: Deterministic contributions

Timing residuals: Difference between maximum-likelihood model and residuals

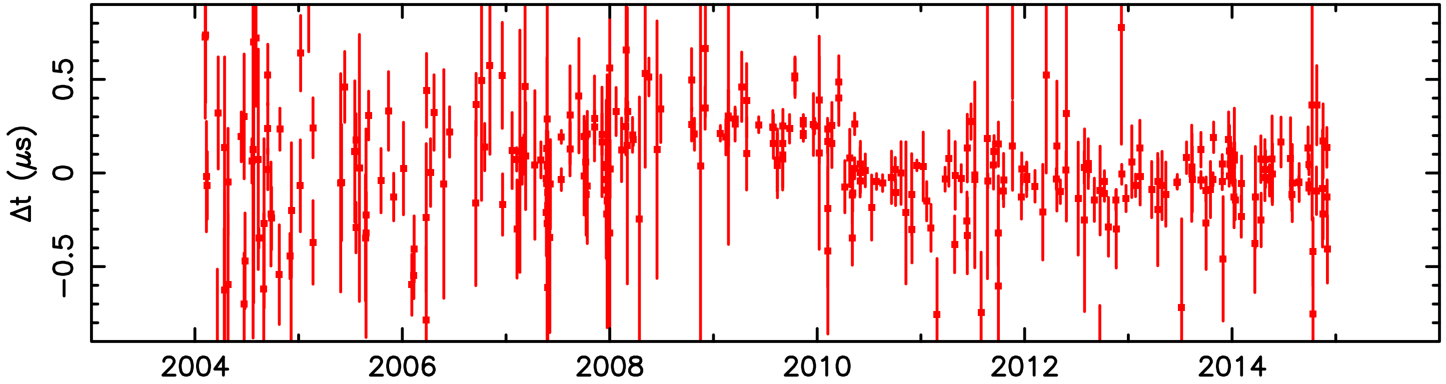
Need to refer arrival times to non-inertial frame: ***Solar System Barycentre.***



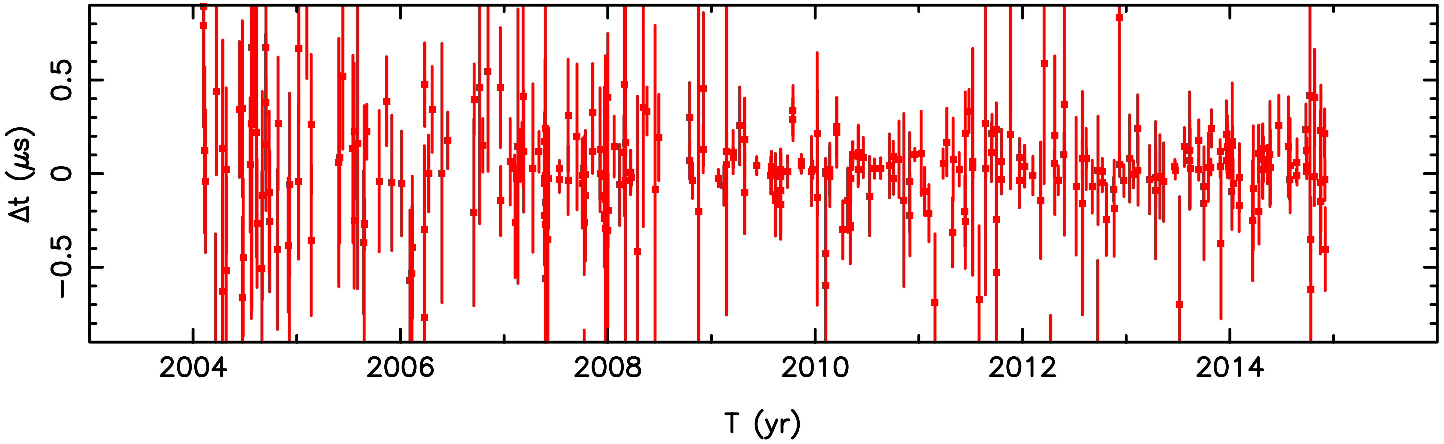
Best model

Timing parallax (curvature of radio waves as it passes solar system)

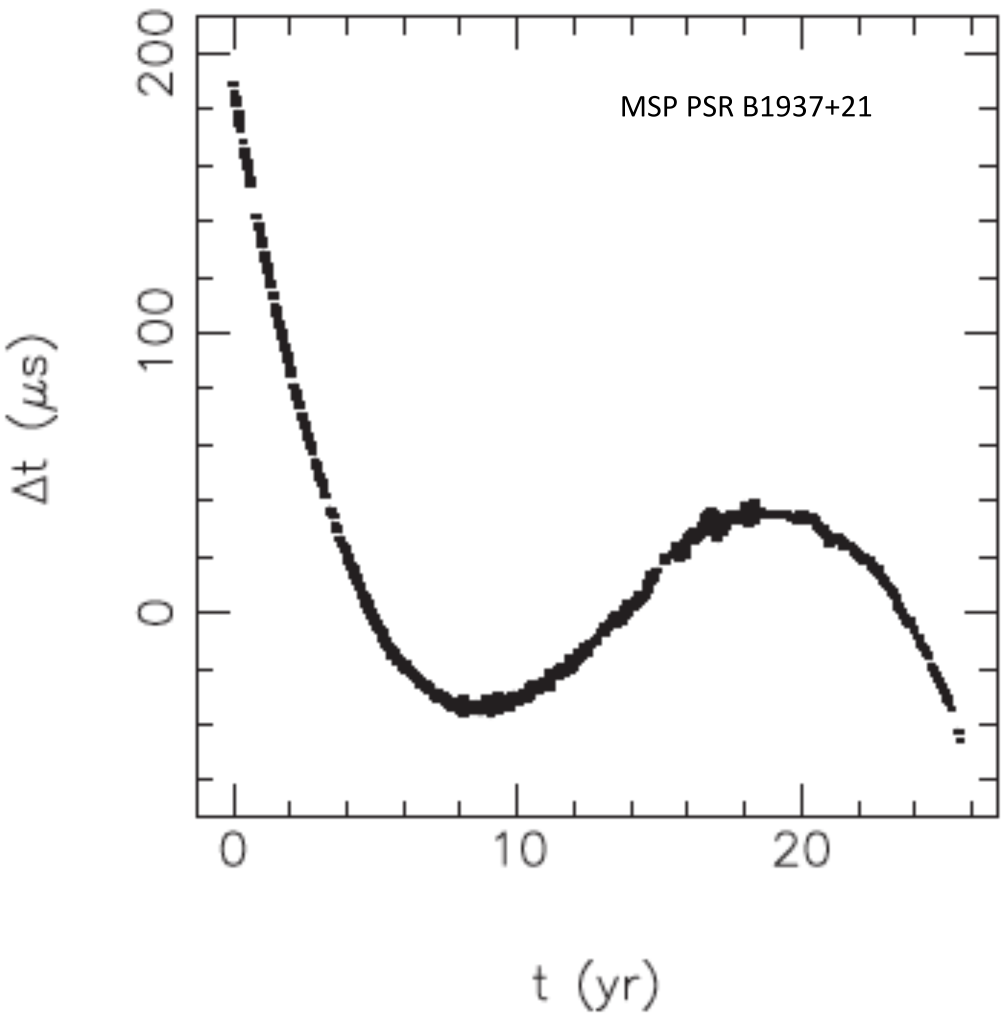
Example: Instrumental offsets



03/05/10 PDFB3 and PDFB4 firmware updated at UT0415 (MJD 55319.18) to remove delays, i.e. to make time-stamp refer to time of signal arrival at digitiser input. In case of PDFB4, the PTU was also changed to remove the two-bin delay. (V6,V3)

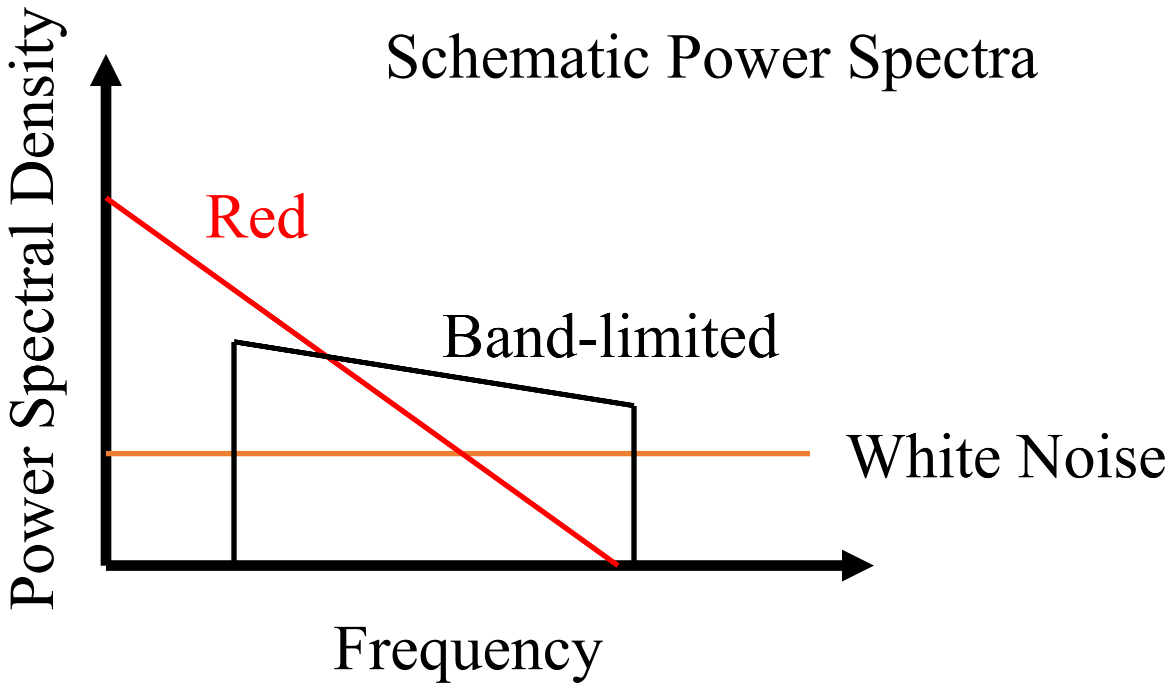


What about this?



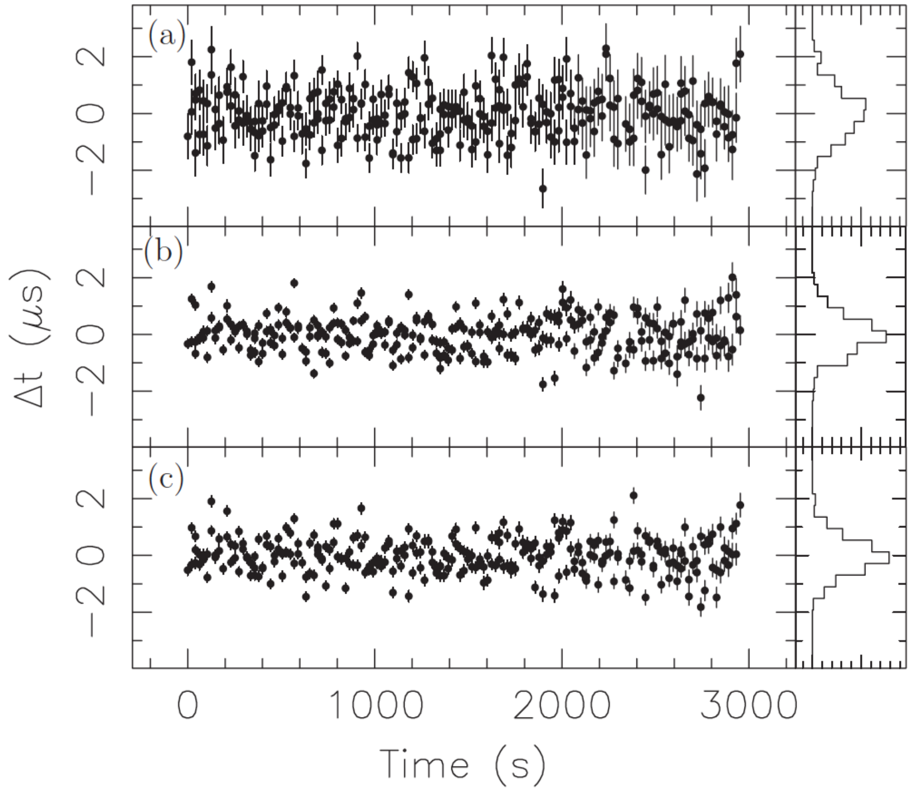
Classes of stochastic processes

- **White Noise (WN)**
  - Uncorrelated
  - Flat power spectrum
- **Red Noise (RN)**
  - Correlated noise
  - Wide-sense stationary
  - Random walks
  - “Red Power spectrum” (more power at lower frequency)
- **Band limited noise**
  - Red power spectrum with cut-offs
  - Excess power at certain frequencies



White noise

- Noise that is uncorrelated on TOA-TOA time scales
- Radiometer noise
- Pulse-shape variations (“jitter”)
- Esoteric ISM effects?
- Instrumental effects?



Pulse-shape variations in PSR J1713+0747 (Arecibo, Shannon & Cordes 2012)

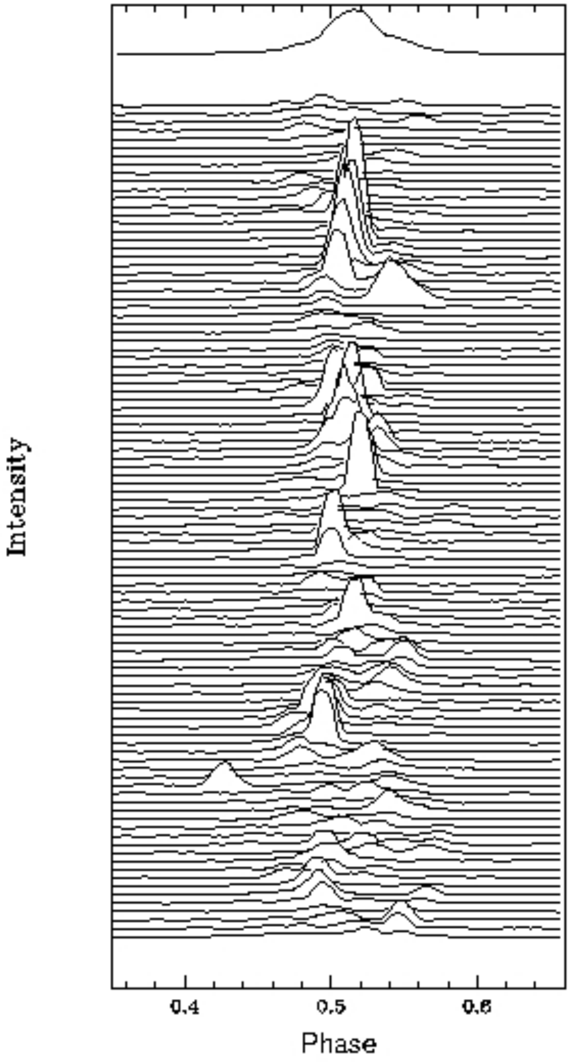
Timing Error from Pulse-Phase Jitter

$$U(\phi) \propto \int d\phi' \, f_\phi(\phi') a(\phi - \phi')$$

$$\Delta t_J = N_i^{-1/2}(1+m_I^2)^{1/2} P \langle \phi^2 \rangle^{1/2}$$

$$= N_i^{-1/2}(1+m_I^2)^{1/2} P \left[\int d\phi \, \phi^2 f_\phi(\phi) \right]^{1/2}$$

- f_φ = PDF of phase variation
- $a(\varphi)$ = individual pulse shape
- N_i = number of independent pulses summed
- m_I = intensity modulation index ≈ 1
- f_J = fraction jitter parameter = ϕ_{rms} / W ≈ 1



Gaussian shaped pulse:

$$\Delta t_J = \frac{f_J W_i (1+m_I^2)^{1/2}}{2(2N_i \ln 2)^{1/2}}$$

$N_6 = N_i / 10^6$

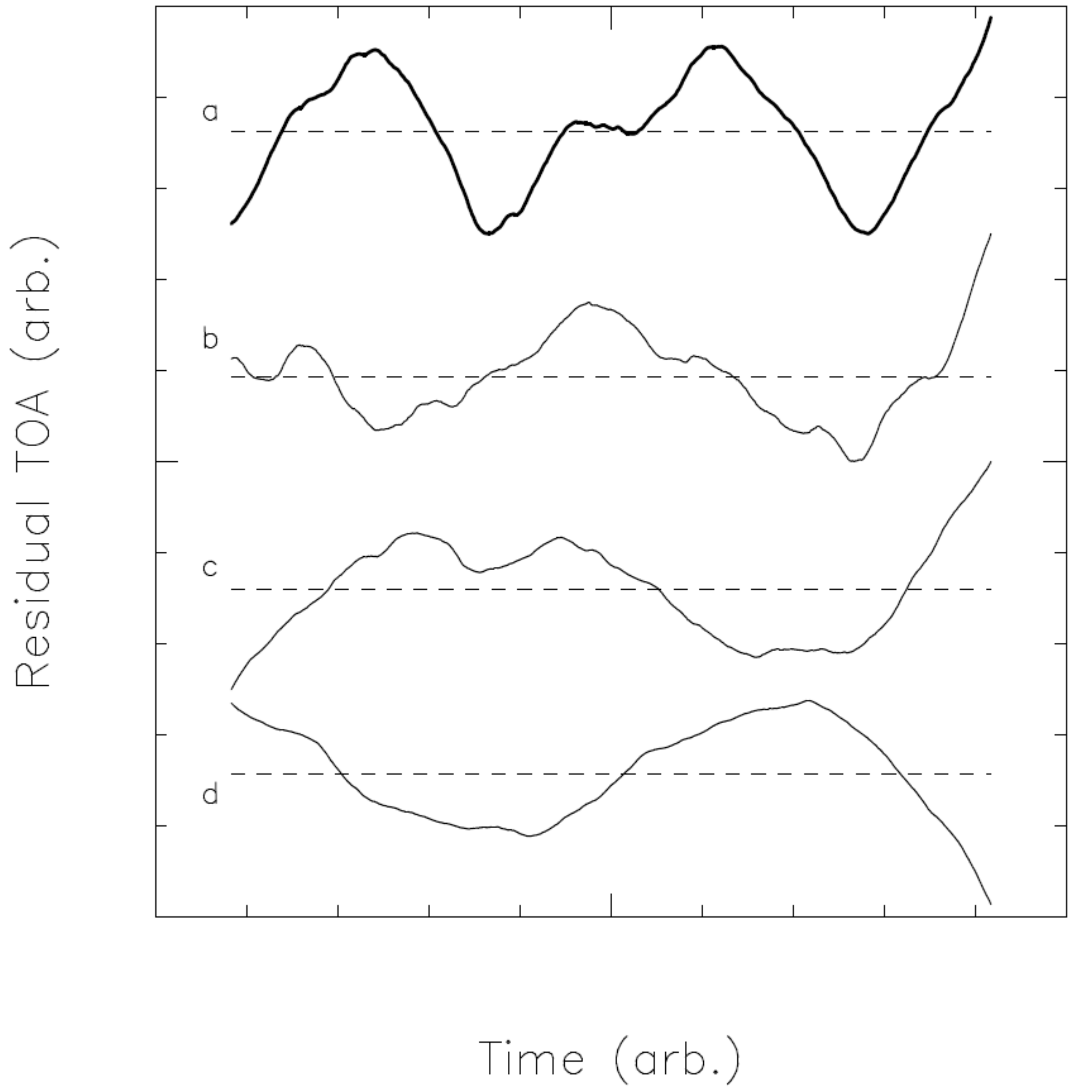
$$\boxed{\Delta t_J = 0.28\mu s \, W_{i,\mathrm{ms}} N_6^{-1/2} \left(\frac{f_J}{1/3}\right) \left(\frac{1+m_I^2}{2}\right)^{1/2}}$$

Red noise

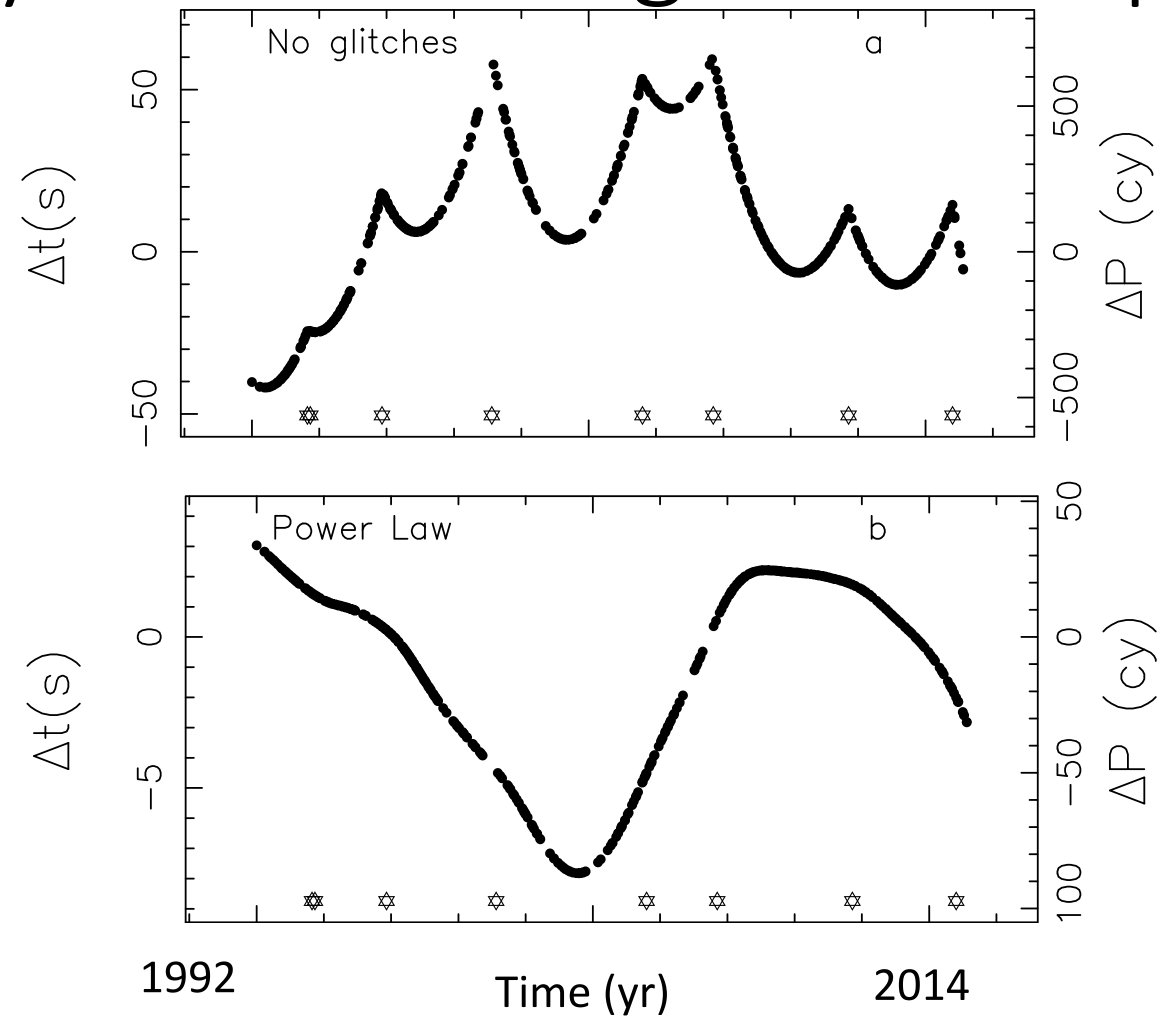
Why don’t we use normal pulsars for PTA work?

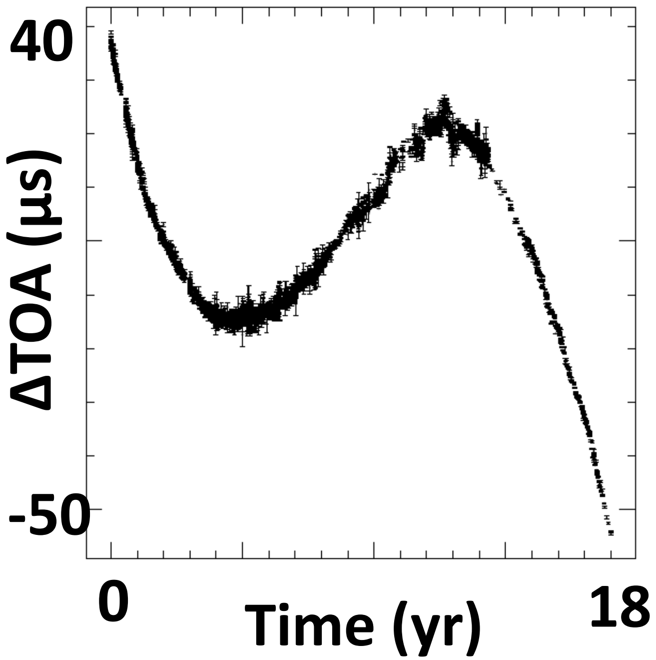
1. Pulse Shape Variations (jitter noise)
2. Timing Noise

Timing noise in an individual pulsar can have very similar shape to gravitational waves



22 years of timing the Vela pulsar





What is causing the systematic variations in residual TOAs?

For these pulsars, the residuals are mostly caused by spin noise in the pulsar:
Torque fluctuations crust quakes superfluid-crust interactions
Other pulsars: excess residuals are caused by orbital motion (planets, WD, NS), ISM variations, GWBs

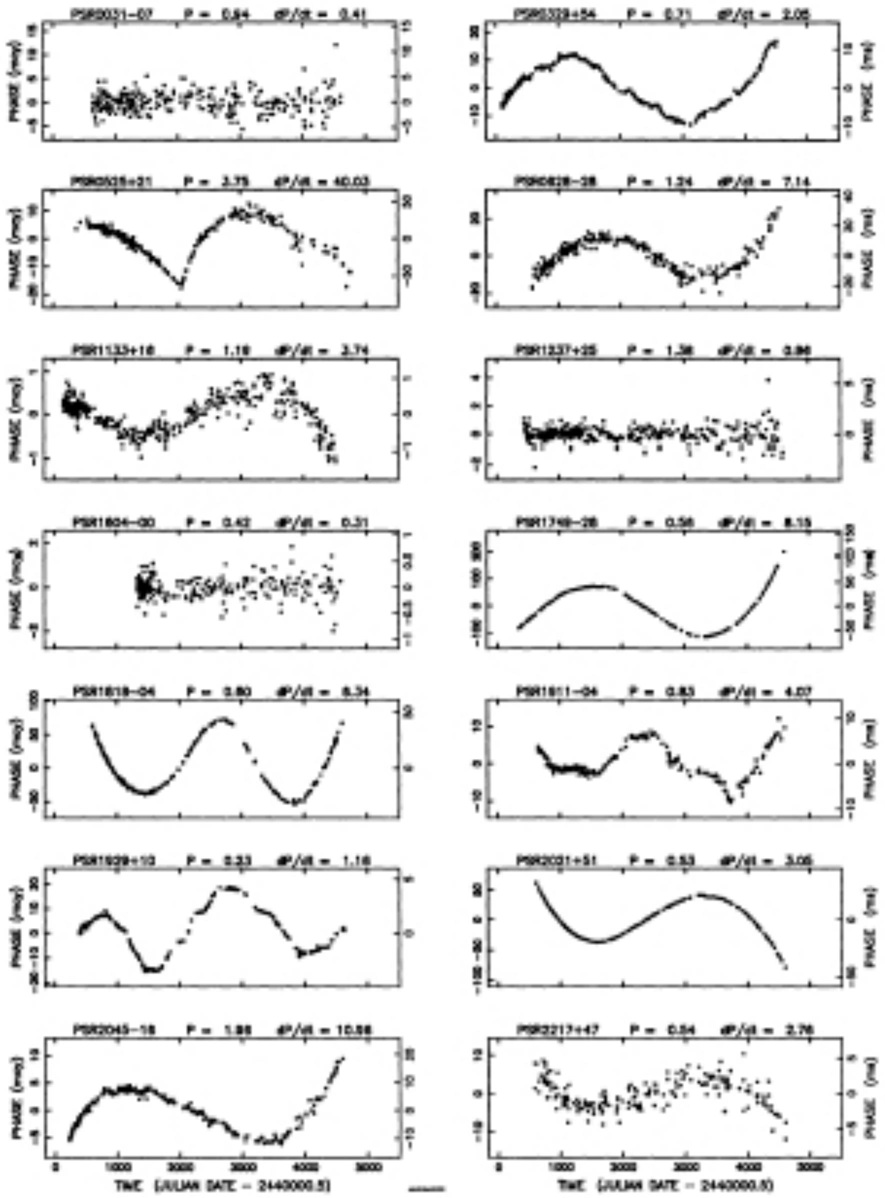
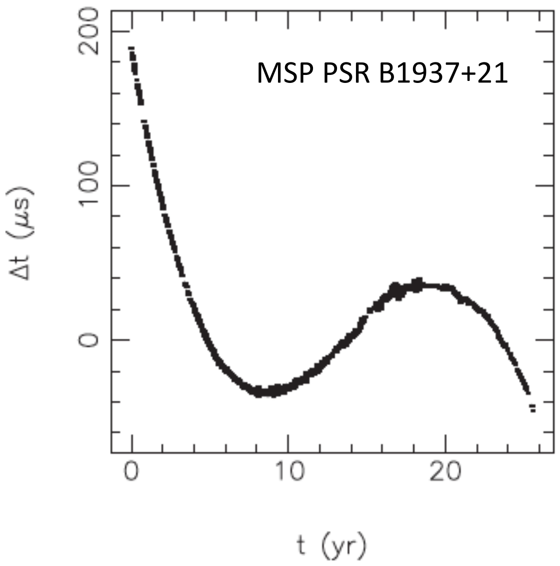
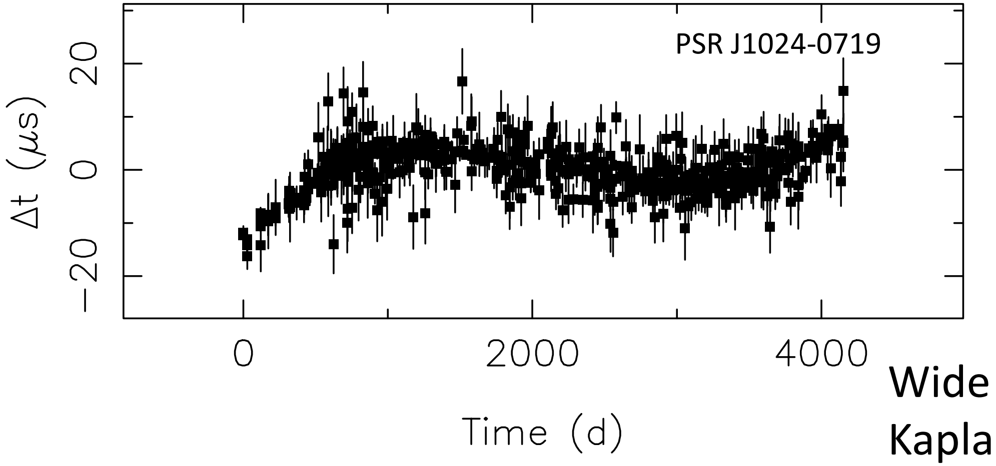


FIGURE I Phase residual curves $\mathcal{R}_2(t)$ for 14 pulsars from the JPL sample of Downs and Reichley (1983). Spin periods P (seconds) and derivatives $\dot{P}$ (in units of 10^{-15} s s^{-1}) are shown at the top of each panel.

Red noise

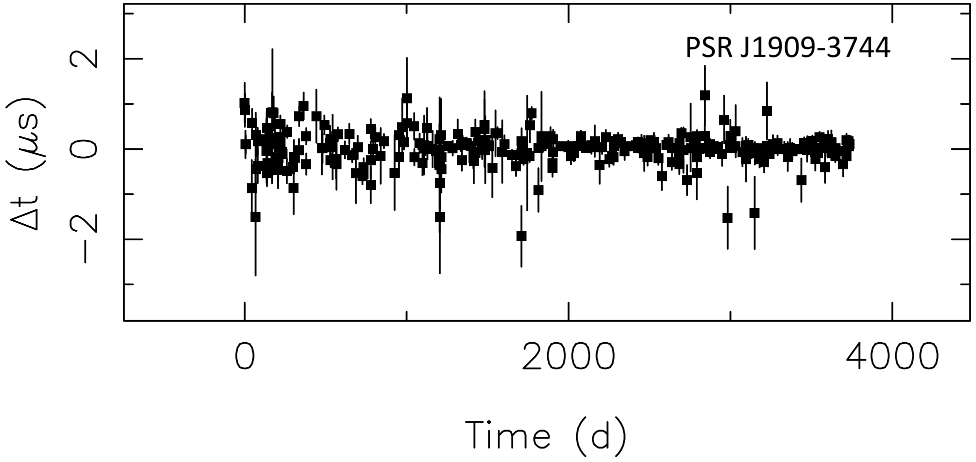
- Most young pulsars show red spin noise
  - Rotation instabilities?
  - Magnetospheric torque changes?
  - Open question: is this a generic property of MSPs too?
  - Can have similar spectral properties to GWB
- Solution: need to (at least) model the presence of red noise in datasets
- Triage bad pulsars



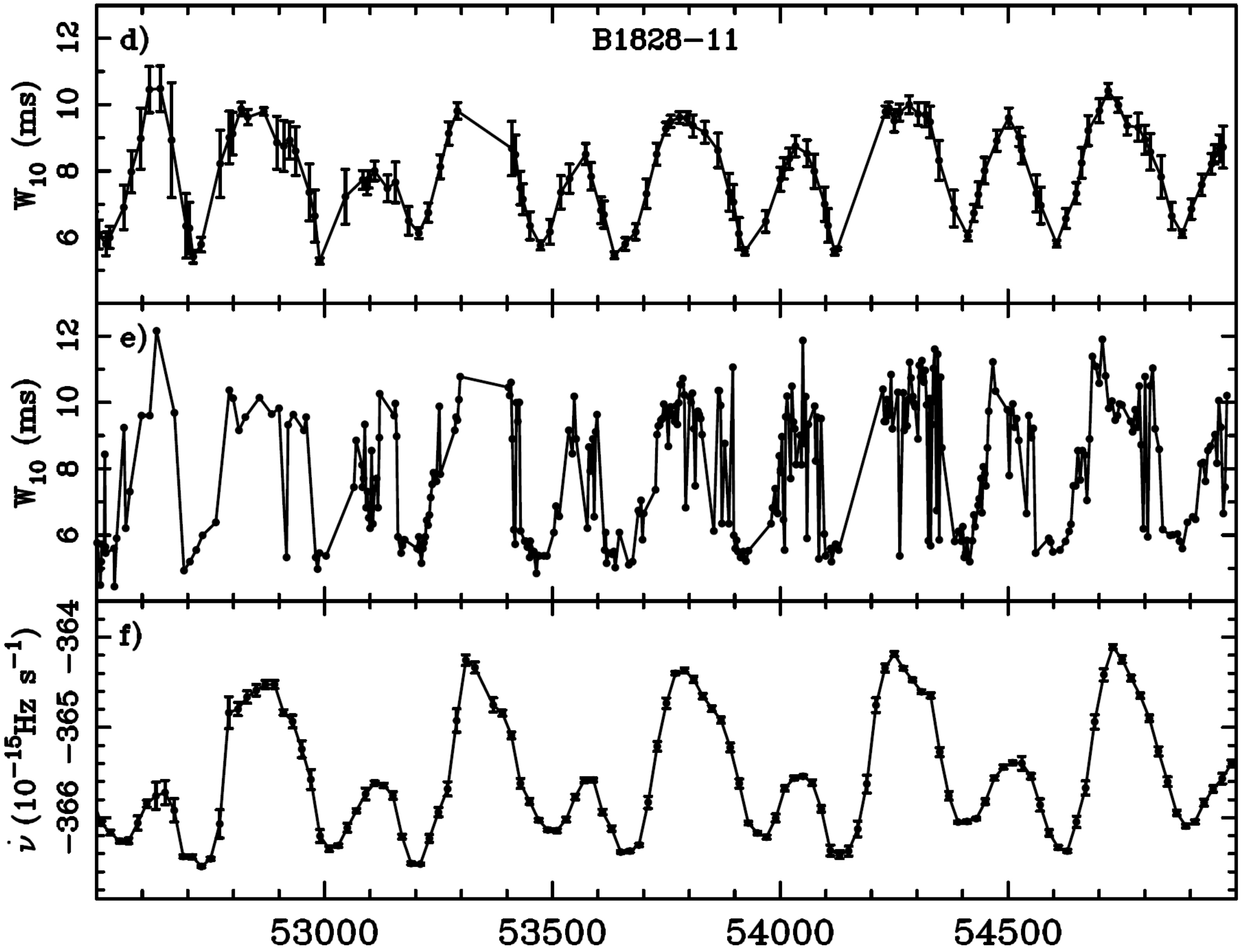


Wide companion
Kaplan et al (2015),
Bassa et al. (2015).

2003 **2014**



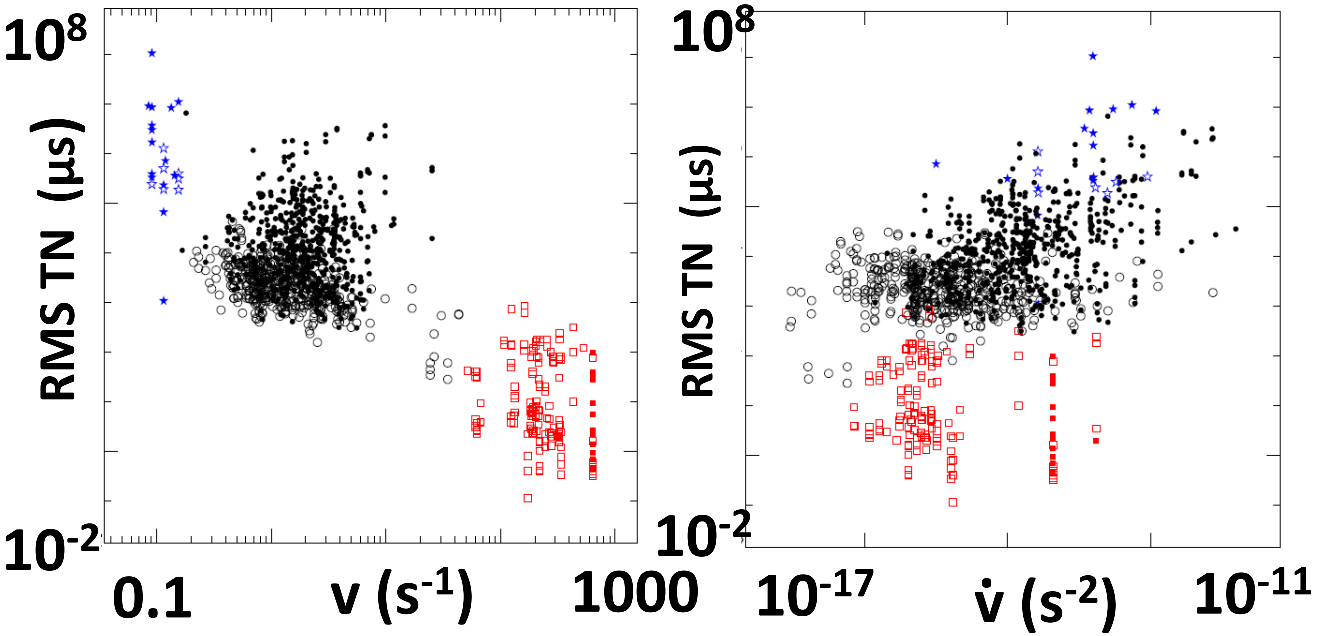
- **Example of quasiperiodic processes (Lyne et al 2010)**



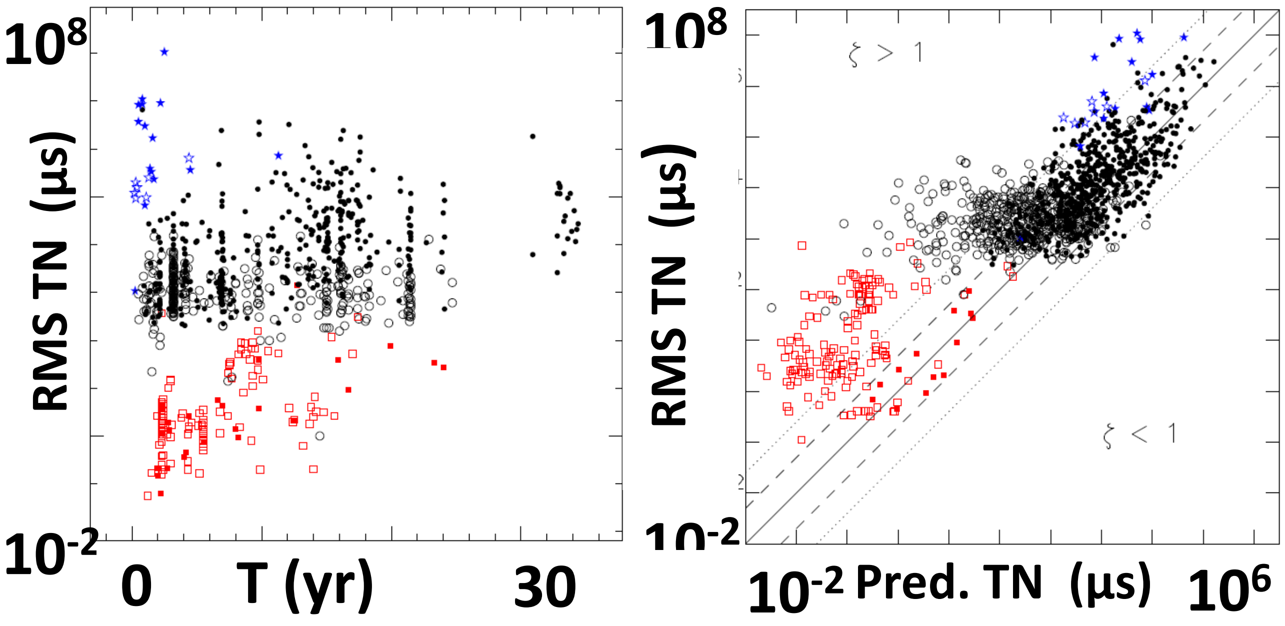
Timing noise across the pulsar population

Examined *every* report of TN in the literature 1980-2010

Blue: Magnetars
Black: Canonical (Normal) Pulsars (CPs)
Red: MSPs
Open symbols: upper limits
Closed symbols: detections



$$\hat{\sigma}_{\mathrm{TN},2} = C_2 \nu^{\alpha} |\dot{\nu}|^{\beta} T^{\gamma}$$ α = -1.4; β = 1.1; γ = 2.0



Implications: Spin noise will be present in MSPs if observed

1. Over longer periods of time
2. With higher timing precision

Fluctuation Properties of Spin Noise + GWB

Spin noise is present in MSPs

Noise has time variability similar to that expected from GWB, making filtering difficult/impossible.

For given spin properties, range of strengths of timing noise.

Going to need to time a larger number of pulsars and then discard the ones that show timing noise.

Plausibly 20 ns (rms) of timing noise over 5 years for typical MSP.

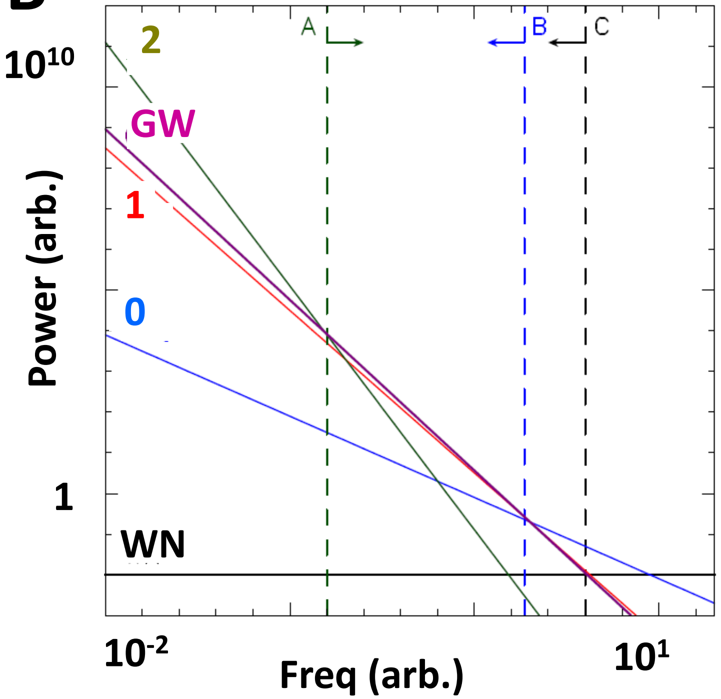


Table 4
Percentage of MSPs Suitable for a PTA

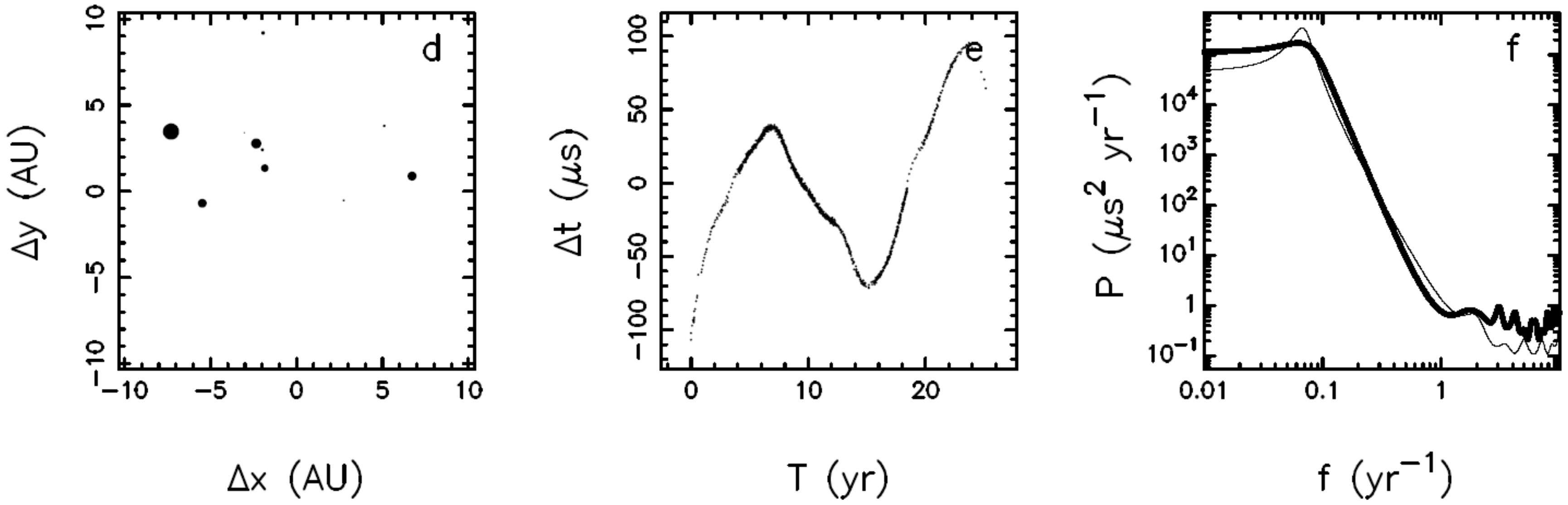
		$N_{\rm PTA}=40$			$N_{\rm PTA}=100$		
T (yr)	$\sigma_{\rm GW,2}$ (ns)	M	$\sigma_{\rm TN,2,t}$ (ns)	$f_{\rm MSP}$ (%)	M	$\sigma_{\rm TN,2,t}$ (ns)	$f_{\rm MSP}$ (%)
2	4	1	2	34 ± 7	1	3	38 ± 8
5	19	3	5	49 ± 8	3	5	52 ± 8
10	59	5	7	58 ± 7	7	8	60 ± 7
20	187	11	10	65 ± 7	13	11	67 ± 7

Asteroid belt Interpretation to Timing Noise in B1937+21

- Low mass circumpulsar system (total mass ~ 0.05 Earth masses)
- 10 -200 objects: Can't resolve periodicities of individual components.

Data tools used

- Simulations of reflex motion
- Remove timing model
- Comparing simulated maximum entropy power spectra to best fit power spectra.



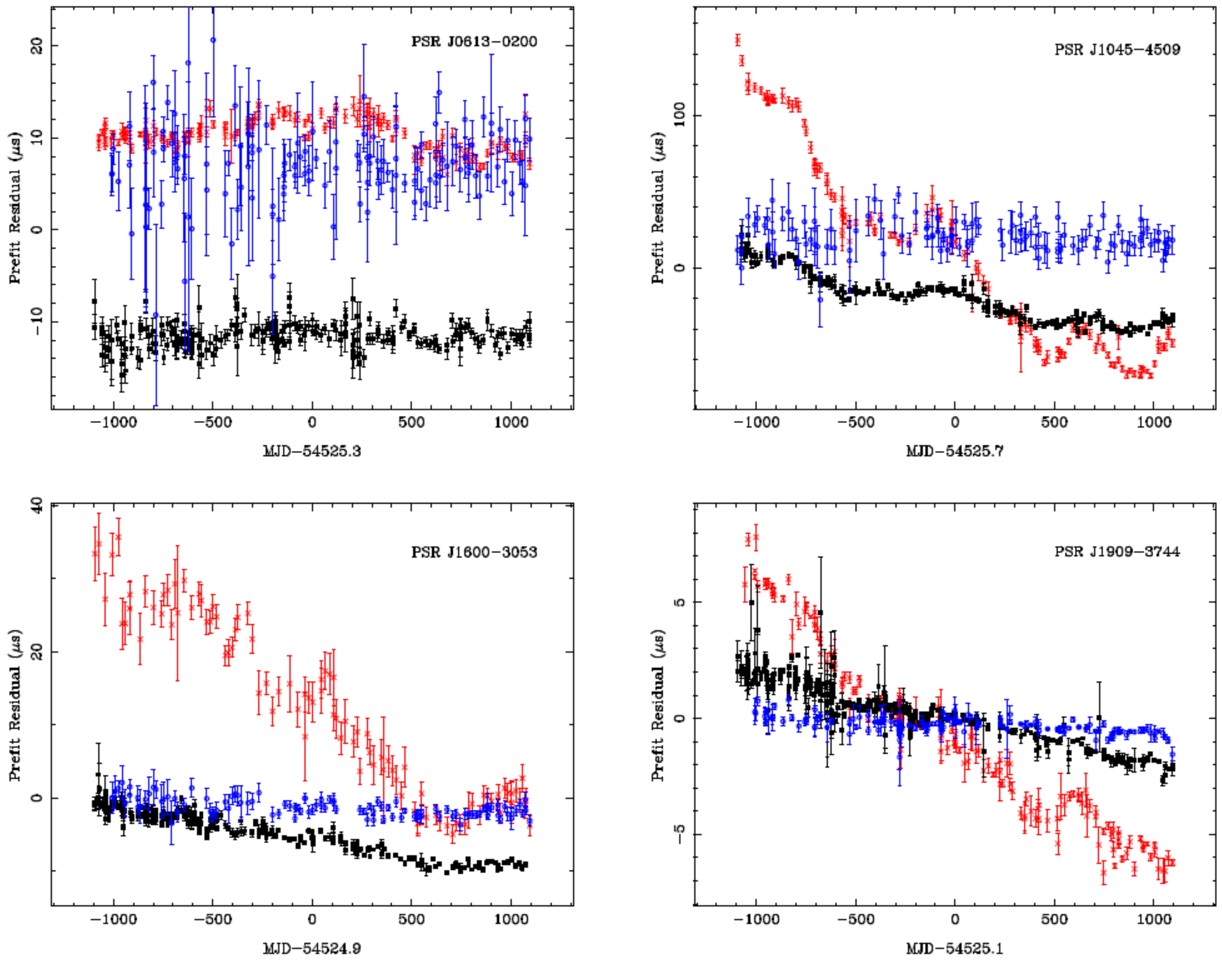
(Shannon et al. 2013)

Red noise from the interstellar medium

Largest red signal in data set: Variations in dispersion measure (DM).
Proportional to λ^2.

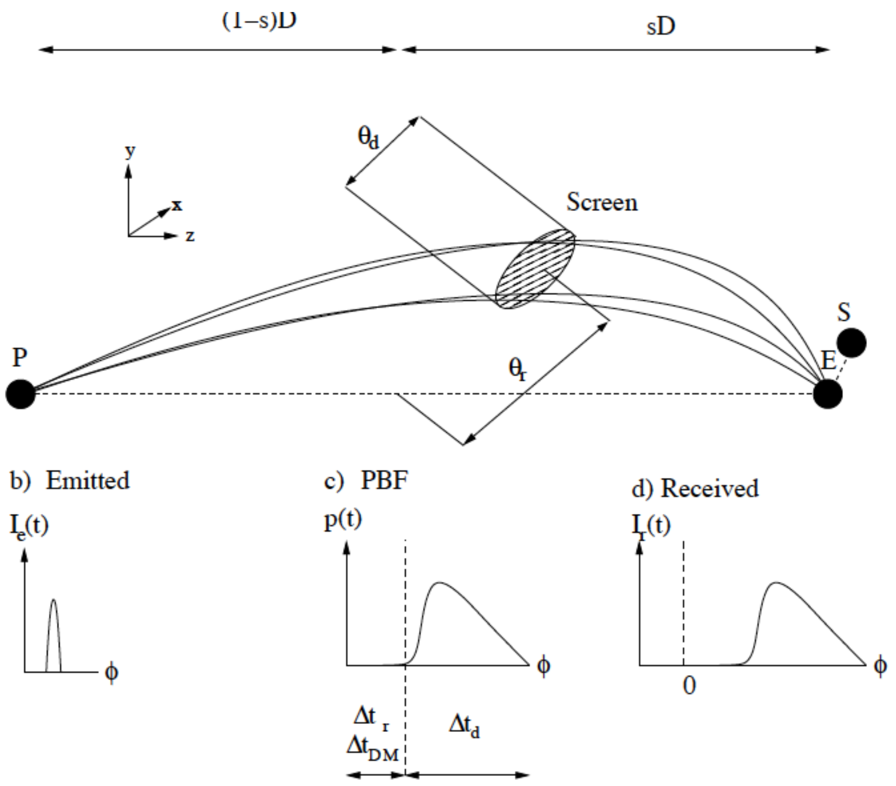
Need to remove **red signal** associated with DM variations without removing red signal associated with GWB

Include model of λ-independent in DM correction algorithm (Shannon 2011, Keith et al. 2013, Lee et al. 2014, Lentati et al. 2014)



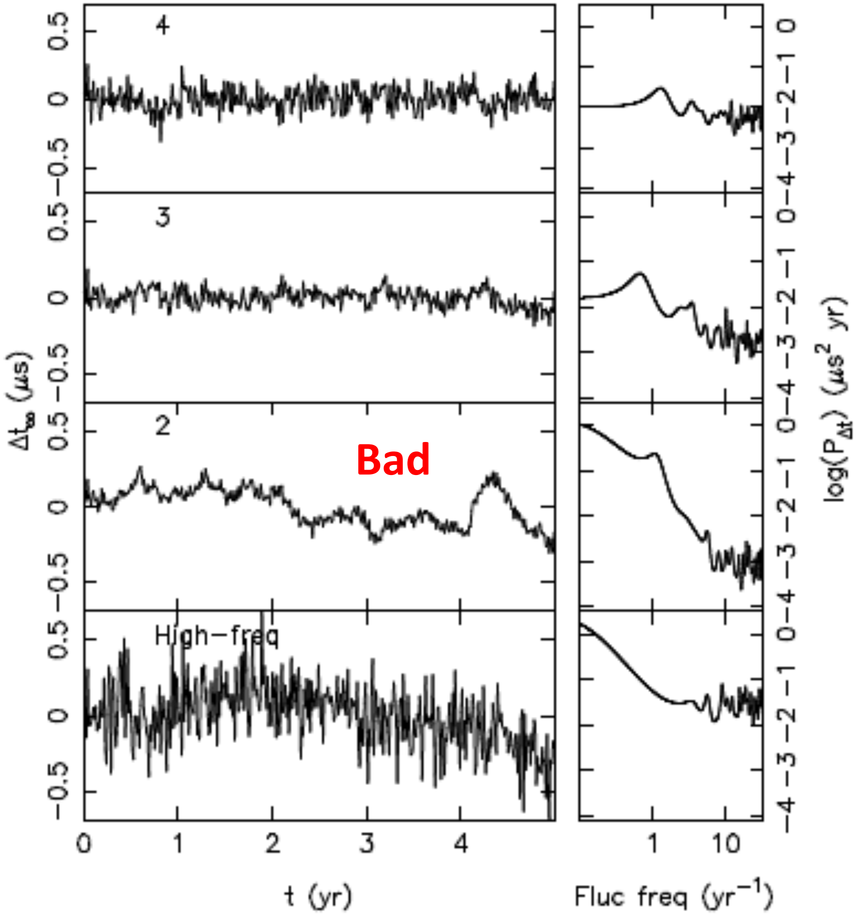
Multi-path propagation

- Multi-path propagation causes broadening of pulse signal.
  - Proportional to λ^2 to $\lambda^{6.4}$
- Broadening is variable with time
- Strongest for distant pulsars observed at low frequencies
- Solution:
  - Observe at higher radio frequency
  - Explore mitigation methods like cyclic spectroscopy (Demorest 2012)



Biases in red-noise estimates

Assumption: observations contain only (instrumental) white noise and DM variations



Correct for DM + Scattering (2)

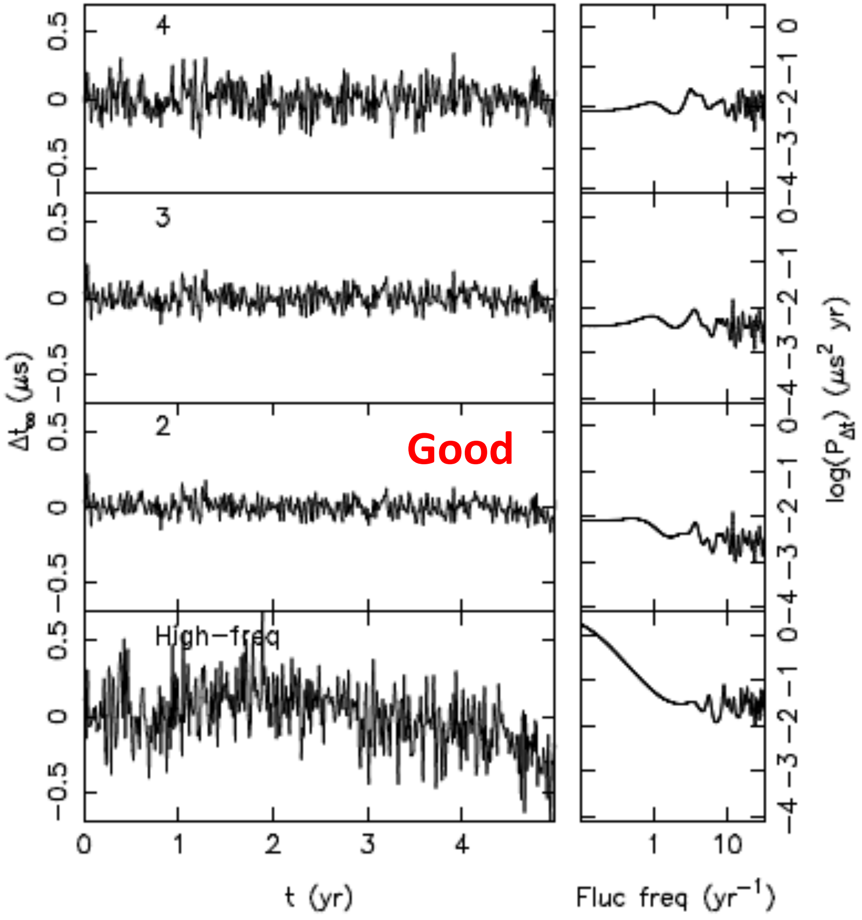
Correct for DM + Scattering

Correct for DM

Only observe at high frequency

Biases in red-noise estimates

Down-weight low frequency TOAs to account for scattering variations.



Correct for DM + Scattering (2)

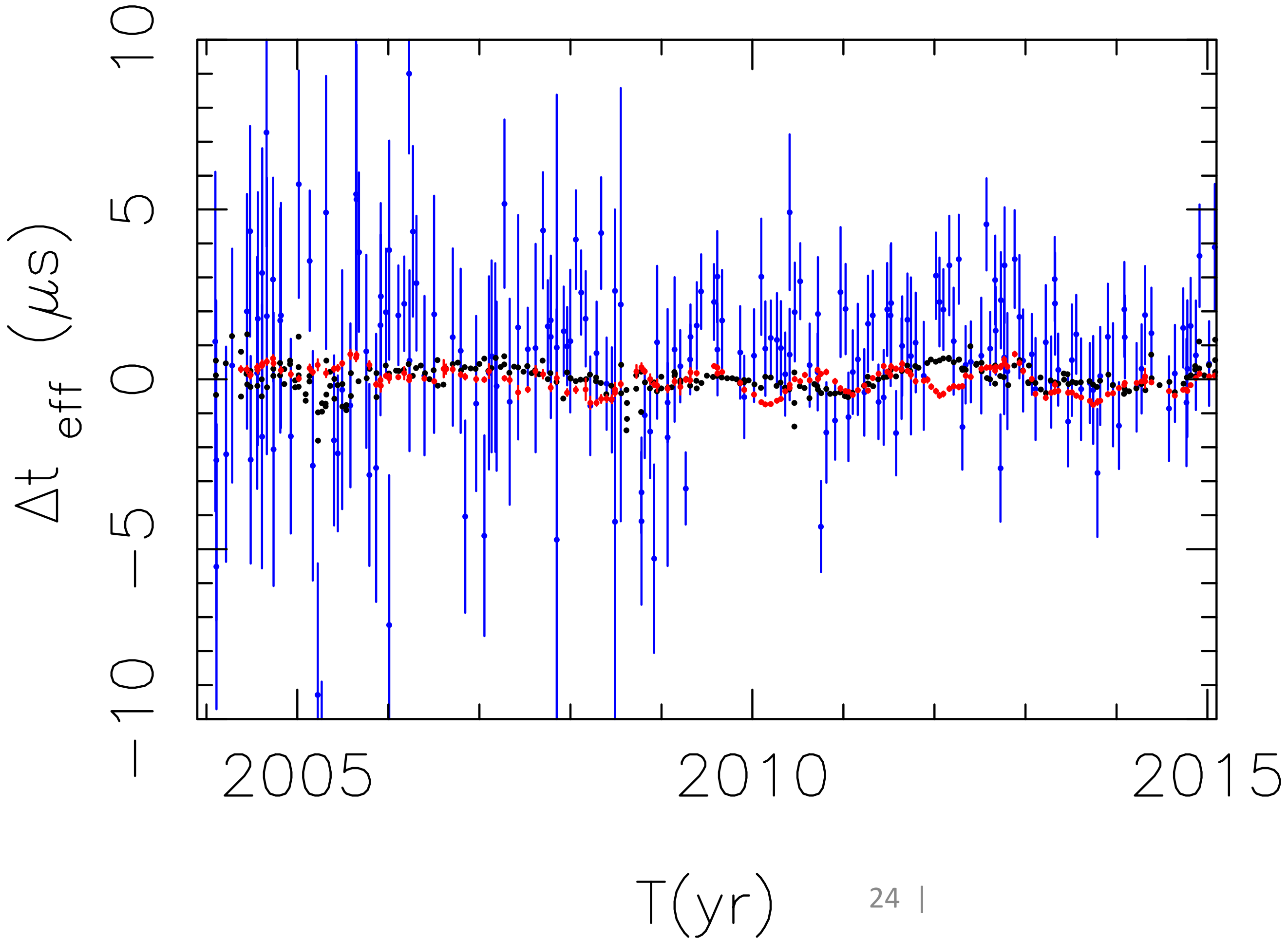
Correct for DM + Scattering

Correct for DM

Only observe at high frequency

Observations of scattering noise

- PSR J1643-1224
- See IPTA Noise paper (Lentati et al. 2016)!



Parkes residuals corrected for DM variations and scaled to 3.1 GHz

Blue: 10cm
Black: 20cm
Red: 50cm

Modelling noise

- **White noise:**
  - Modification to TOA uncertainties: EFAC and EQUAD
- **Red noise:** power spectrum
  - Useful for wide-sense stationary processes
  - Maps onto gravitational wave background
  - Power law: $P_r(f) = A\, f^{-\beta}$
    - In Bayesian methods A and β can be directly included in model
  - More complicated models appropriate for band-limited processes
- **Band noise**
- **System noise**
- **Non-stationary models**

What's the point of spectral estimation?

- Searching for deterministic signals: periodic
- Stochastic signals: described by amount of signal per frequency
  - Useful way of expressing properties of time-correlated signals.
- Describe power spectral density (PSD). Integral under power spectral density estimator (PSDE) is defined to be the variance in the time series.

Example: Fourier Transforms

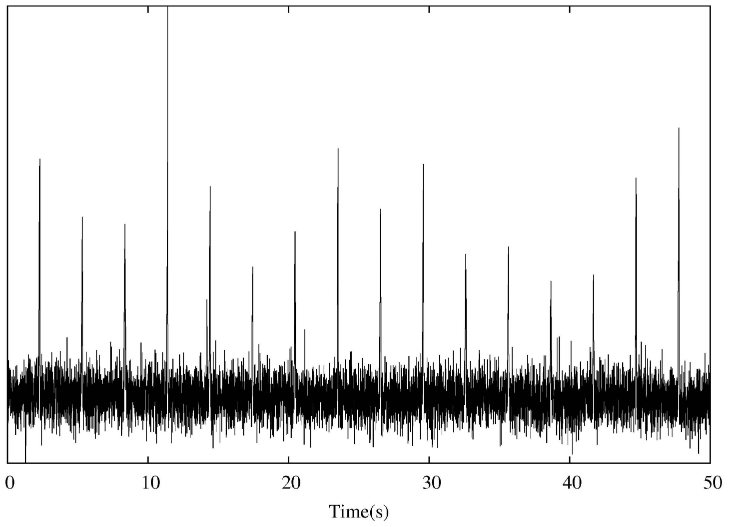
- Advantage (can do it “fast”) useful if doing a lot of computation
  - Very useful for online computation

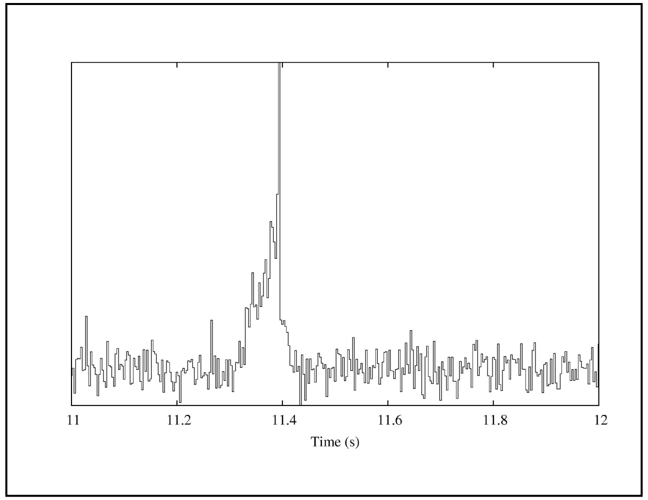
- Beware of **spectral leakage.** If you measure P(f) ~ f^{-2} it is likely due to spectral leakage

Example: Baseband data

- Voltage data recorded as 2-bit numbers with VLBI recorder

- Convert data to intensity as a function of frequency
  - Turn data into power spectra: $|FFT|^2$

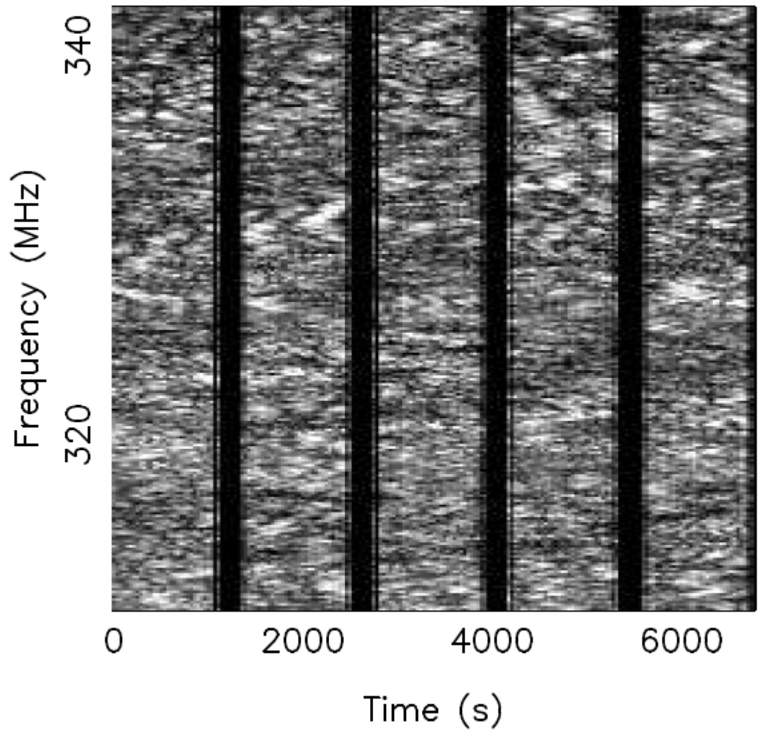


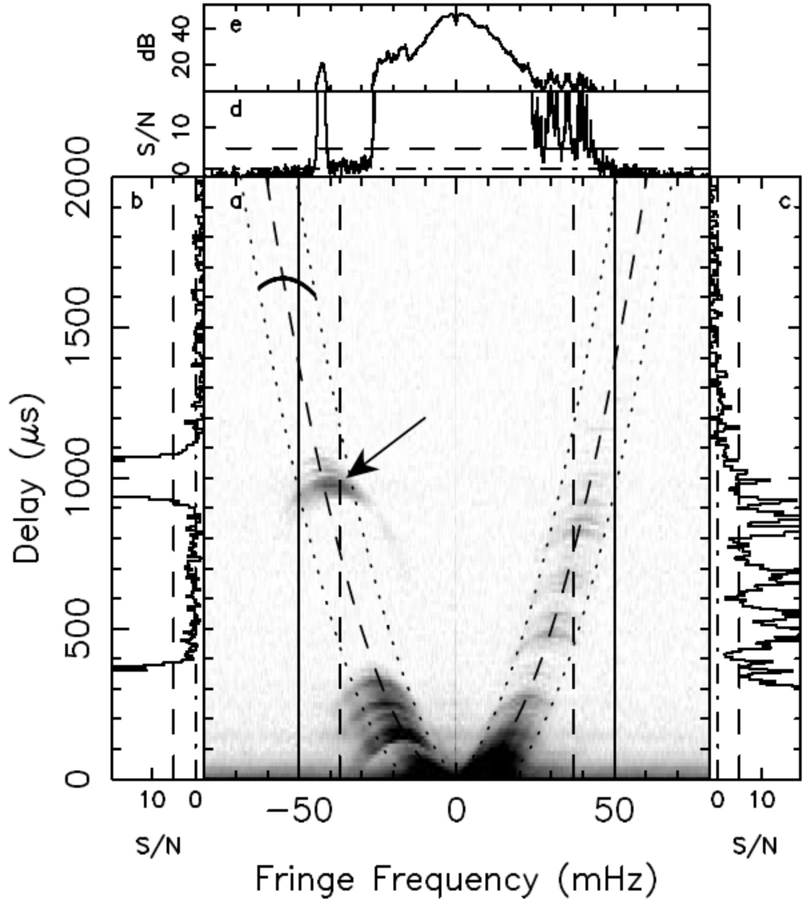


Dynamic and secondary spectra

- Look at intensity as a function of frequency
    - Dynamic spectrum
    - Stochastic pattern

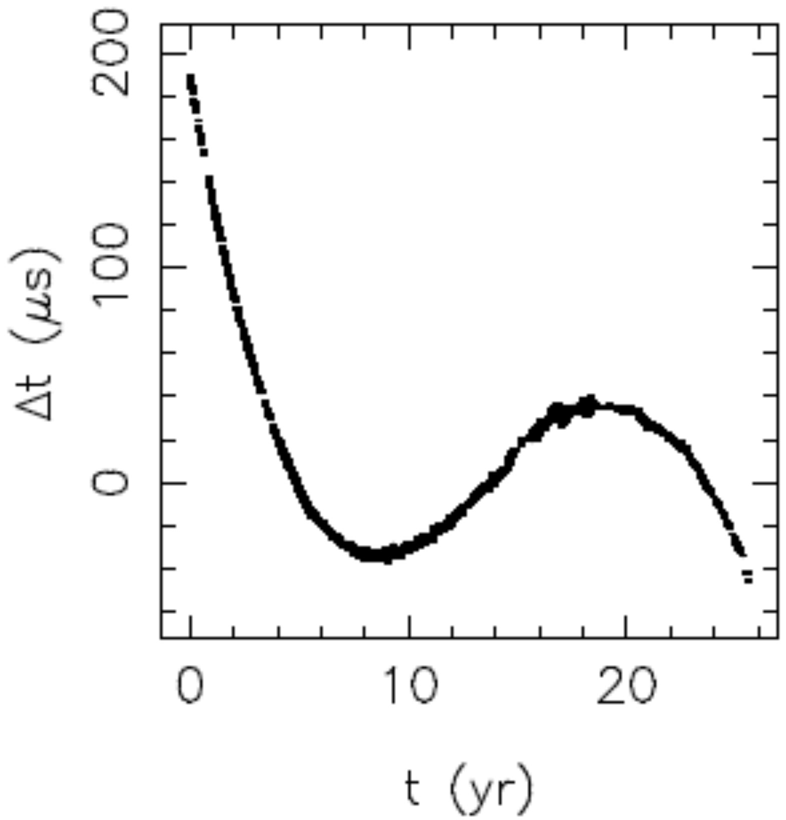
- Second FFT
    - Structure is more isolated

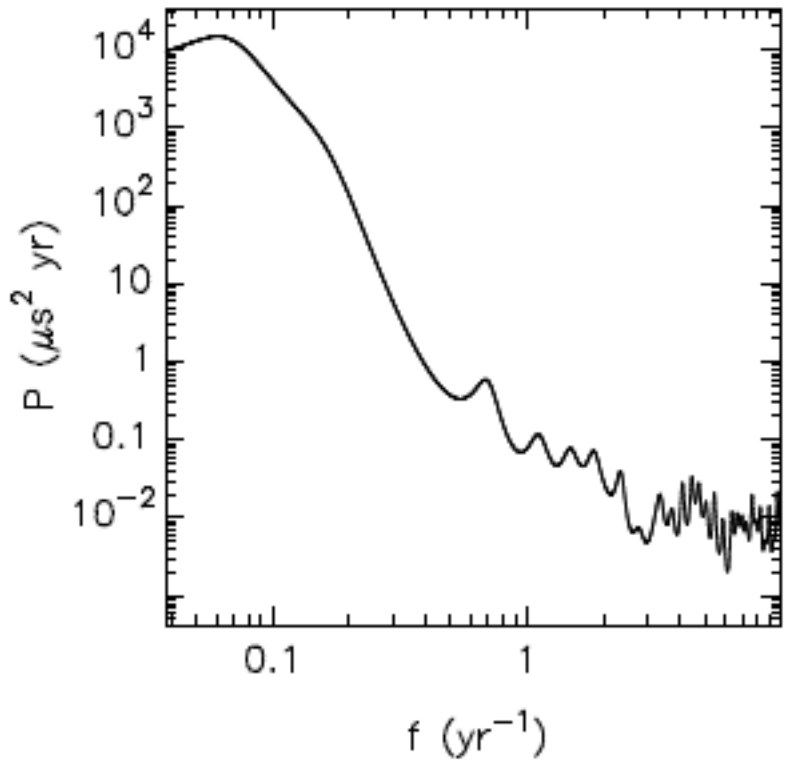




Maximum entropy spectra estimation

- Fourier approach: model autocorrelation function, set to zero outside spectral window
- Define process to have *maximum entropy* (fewest assumptions about data, no sharp edges outside of data span)





Cholesky spectral estimation

- If you have a good guess of what the statistics of your stochastic signal, then you should use them as part of your fitting process

- Lomb-Scargle: fitting sinusoids to irregularly sampled data

- Weighted Lomb Scargle: Assume noise covariance matrix is diagonal (means samples are assumed to be statistically independent).

- Weighted Lomb-Scargle with red noise covariance matrix
    - Cholesky spectral estimation implemented as spectralModel plugin in tempo2

Autocorrelation and cross correlation functions

$$(f \star g)[n] \stackrel{\mathrm{def}}{=} \sum_{m=-\infty}^{\infty} f^*[m]\ g[n+m].$$

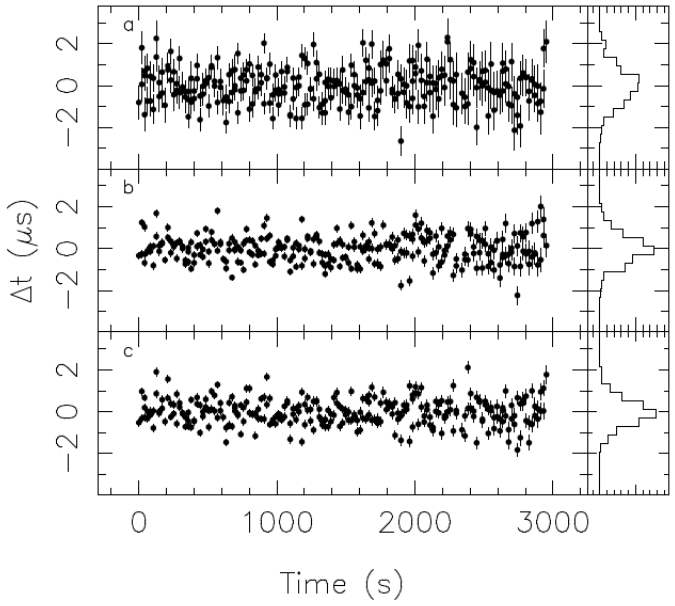
- Useful for measuring characteristic size (timescale etc.) for stochastic process
  - Covariance versus lag (time, frequency)

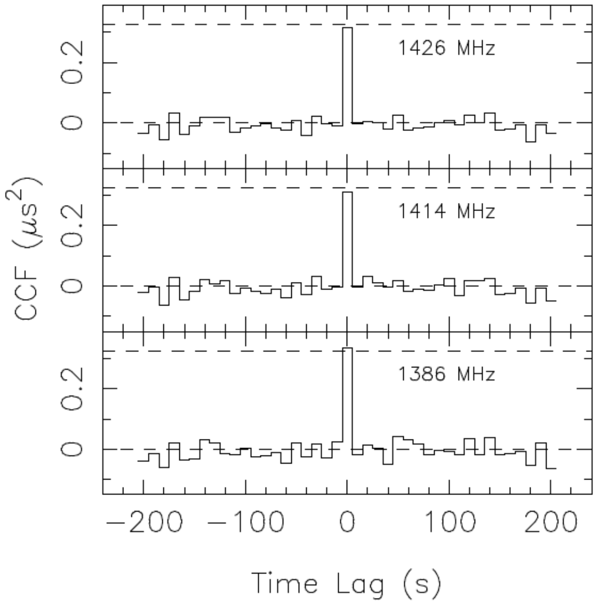
- Correlation length for white signal ~ sampling cadence (spike at zero lag)

- Correlation length for red signal is ~ T_{obs}

Identifying effects of pulse shape variations

- Make TOAs in multiple sub-bands
- Calculate cross-correlation functions of residuals TOAs
- Example: Arecibo observations of J1713+0747





Simulating red noise

- Lots of ways to do it!

- Using a Fourier transform:
  - Simulate complex random samples
  - Multiply complex variables by bandpass
  - Invert using FFT
  - Issue: need to simulate band pass frequency 1/T to avoid edge effect
    - (FT makes assumption that signal is periodic).
  - Make data Hermitian so that time series is Real

- Other methods
  - Sum of oscillators (used in tempo2 and IPTA data challenge)
  - Use covariance function + Cholesky decomposition

Summary: physical model for timing effects

Table 1

Selected Timing Effects

Term	Type[a]	Mean Part: Symbol	Mean Part: Value	Stochastic Part: Symbol	Stochastic Part: Value	Achromatic or Chromatic[b]	Fluctuation Spectrum: Signature[c]	Fluctuation Spectrum: Shape	PSR-PSR Correlation[d]	Comments
Spin rate	A	$t_{\rm spin}$	yr	$\Delta t_{\rm spin}$	$\mu s - s$	a	B, R	$f^{-4} - f^{-6}$	U	
Magnetosphere:										
Pulse Shape	A, T	$t_{\rm P}$	μs − ms	—	—	c	—	—	U	$\nu^{-0.3}$
Pulse Jitter	A, T	—	—	$\Delta t_{\rm J}$	$< \mu$s − ms	c	W, B	see text	U	$\nu^{-0.3}$
Orbital	A	$t_{\rm orb}$	hr	$\Delta t_{\rm orb}$	$<$ ms	a	L, R	$f^{-5/3}$	U	
Dispersion	A, T	$t_{\rm DM}$	$\lesssim s$	$\Delta t_{\rm DM}$	$\lesssim 100\mu$s	C	R	$f^{-5/3}$	U	ν^{-2}
Faraday Rotation	A, T	$t_{\rm RM}$	$\lesssim \mu s$	$\Delta t_{\rm RM}$	$\lesssim ns$	C	R	$f^{-5/3}$	U	ν^{-3}
Interstellar Turbulence										
Pulse Broadening	A, T	$t_{\rm PBF}$	$ns - s$	$\Delta t_{\rm PBF}$	$<$ ns − ms	C	—	complex	U	$\nu^{-4.4}$
DISS	A, T	—	—	$\Delta t_{\delta\rm PBF}$	$\lesssim \mu s$	C	W	flat	U	$\nu^{-1.6}$ - $\nu^{-4.4}$
RISS	A, T	$t_{\rm PBF,RISS}$	$\lesssim \mu s$	$\Delta t_{\delta\rm PBF,RISS}$	$\lesssim \mu s$	C	R	$f^{-7/3}$	U	?
Angle of Arrival	A, T	—	—	$\Delta t_{\rm AOA}$	$\lesssim \mu s$	C	R	$f^{-2/3}$	U	ν^{-4}
Angle of Arrival	A, T	—	—	$\Delta t_{\rm AOA,SSBC}$	$\lesssim \mu s$	C	R	$f^{-1/3}$	U	ν^{-2}
Multipath averaging	A, T	—	—	$\Delta t_{{\rm DM},\nu}$	$\lesssim 0.1\mu$s	C	R	complex	U	$\nu^{-23/6}$
Astrometric[e]	T	$t_{\rm AST}$		$\Delta t_{\rm AST}$		a	—	—	U	
Newtonian solar perturbations	T	—	—	$\Delta t_{\rm Newt,SSBC}$		a	—	—	C	dipolar
Radiometer Noise	T	—	—	$\Delta t_{\rm S/N}$	$< \mu$s − ms	c→C	W	flat	U	$\nu^{0} \to \nu^{-2.7}$
Polarization	T	—		$\Delta t_{\rm pol}$		c	W	flat	U	
Gravitational Lensing	A	$t_{\rm GL}$		$\Delta t_{\rm GL}$		a	—	—	U	Episodic
Cosmic Strings	A	$t_{\rm STR}$		—		a	R	$f^{-16/3}$	U	Red noise if multiple events
Gravitational Waves	A	—	—	$\Delta t_{\rm GW}$	$\lesssim 100$ ns	a	R	$f^{-13/3}$	C, U	Two terms

Cordes & Shannon

[a] A = astrophysical, T= timing estimation error

[b] a = achromatic, C = strongly chromatic, c = weakly chromatic

[c] Fluctuation spectrum properties: R = red, W = white, B = bandpass, L = lowpass

[d] U = uncorrelated between different pulsar lines of sight, C = correlated

[e] Includes clock errors and Earth spin variations

Cordes & Shannon (2010, arXiv:1010.3785)

Signal model for pulsar timing

- TOAs = timing model + white noise + red noise

- Timing model = deterministic terms

- Red noise = **(gravitational waves)** + spin noise + ISM noise + …

- White noise = radiometer noise + pulse jitter + instrumental effects

- Can fit/marginalize deterministic terms without too much fuss
- Want to minimize stochastic contributions relative to GWs
- Need to incorporate stochastic contributions into Likelihood function

Maximum-likelihood methods

- If observations are normally distributed, can calculate the probability (p) that the parameters agree with the data:

$$p(d|\lambda) = \frac{1}{\sqrt{2\pi\sigma^2}} \prod_i \exp\left(- \frac{(d_i - s_i)^2}{2\sigma^2} \right).$$

  - d_i = data points
  - s_i = model at data points
  - σ = uncertainty in model

  - Strategy 1: find parameters that maximize probability (usually minimize –log(p)="χ^2"

Maximum-likelihood methods

- If parameters are linear, then the problem is linear, and parameters (and uncertainties) can be solved for using a set of linear equations
  - Convergence in single iteration of fit
  - There is a unique global minimum

- If model is non-linear it is often possible to linearize the problem about initial parameters
  - Binary orbital parameters are very non-linear
  - Convergence may take a few iterations if close to minimum
  - Not guaranteed to find global minimum

- Downsides:
- Difficult quantify the effect of assumptions in the uncertainties
- Minimum in χ^2 not representative if PDF is misshapen

Strategy 2: Bayesian analysis

- Want to calculate probability of Model ***M*** with parameters ***X*** given data ***D***

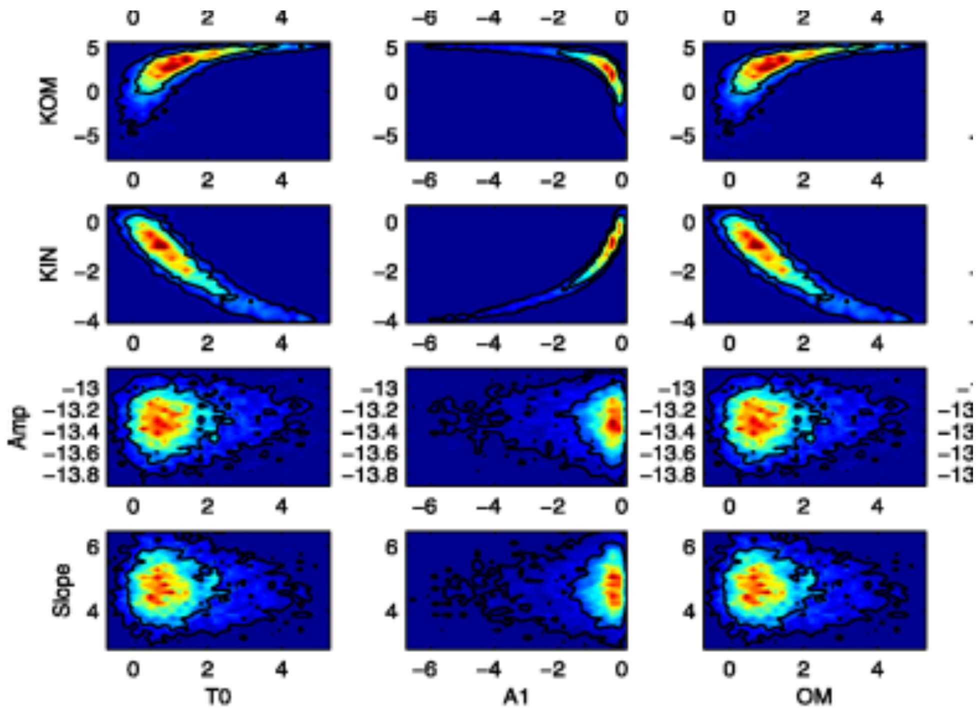
- Bayes Theorem:
  $p(\boldsymbol{X}|\boldsymbol{D},\boldsymbol{M}) = p(\boldsymbol{D}|\boldsymbol{X},\boldsymbol{M})P(\boldsymbol{X}|\boldsymbol{M})/P(\boldsymbol{D}|\boldsymbol{M})$

- **Posterior probability:** $P(\boldsymbol{X}|\boldsymbol{D},\boldsymbol{M})$ (the question you want to answer)
- **Likelihood:** $P(\boldsymbol{D}|\boldsymbol{X},\boldsymbol{M})$ (How well does you data match the parameters and model)
- **Prior:** $P(\boldsymbol{X}|\boldsymbol{M})$ (The range of expected values for the parameters)
  - Usually choose flat (uniformed priors), but can also choose physical priors
- **Evidence:** $P(\boldsymbol{D}|\boldsymbol{M})$ (How well the model matches the data)

Strategy 2: Bayesian analysis

- Instead of finding the most likely set of parameters we consider parameters (or sets of parameters) of interest and **marginalize** over the rest

- Complete picture for the probability distribution for a given parameter



Calculating p($X|D,M$)

- Challenge: for parameter estimation, want to marginalize over parameters that you don't care about (or care about independently)

- For Gaussian distributions with flat priors this can be calculated analytically

- Need creative (numerical) way to sample p($\boldsymbol{X}|\boldsymbol{D},\boldsymbol{M}$) for complicated PDFs.

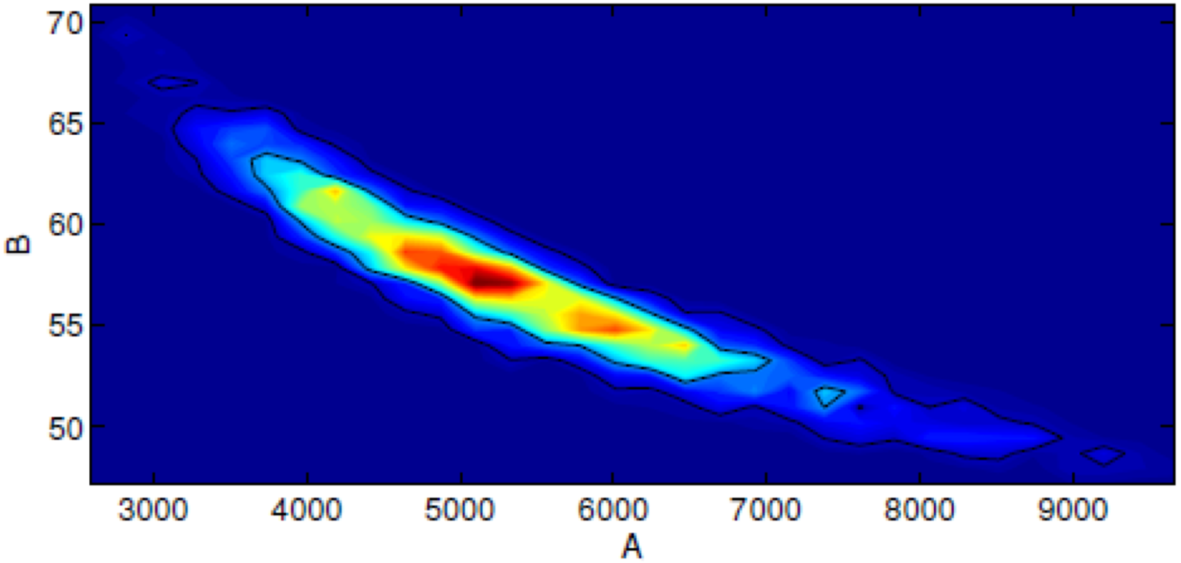
- Various algorithms:
  - Markov-chain Monte Carlo (MCMC), Nested Sampling, Polychord, etc.

Marginalization

- Key to Bayesian analysis: Integrate over ‘nuisance’ parameters, characterise the parameters that you are interested in

- Things you don’t care about but that affect the answer you want

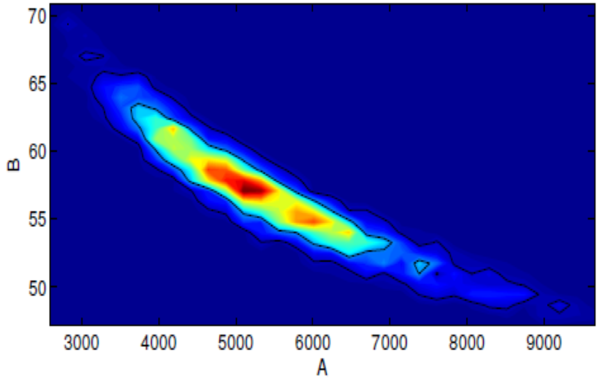
- Consider 2-dimensional problem –
- Probability density for parameters
- A and B.

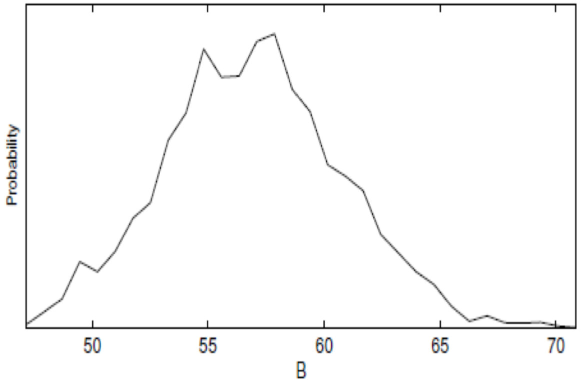


Marginalization

- Can marginalise numerically after sampling

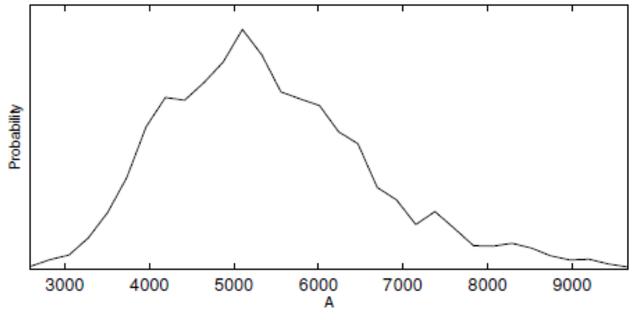
Integrate over A to get the probability of B





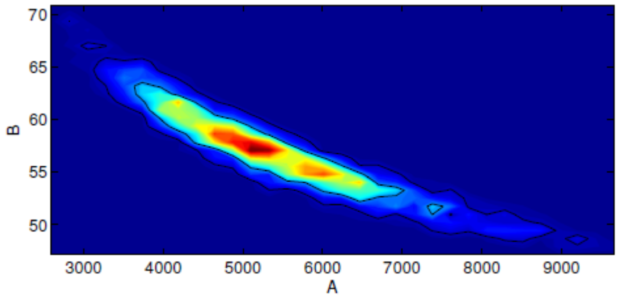
Can marginalize numerically after sampling

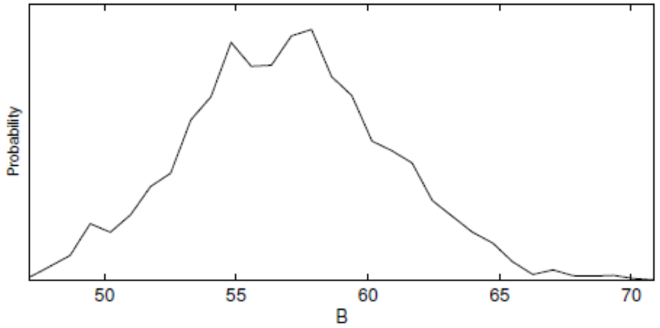
- Can marginalise numerically after sampling



Can also marginalise analytically

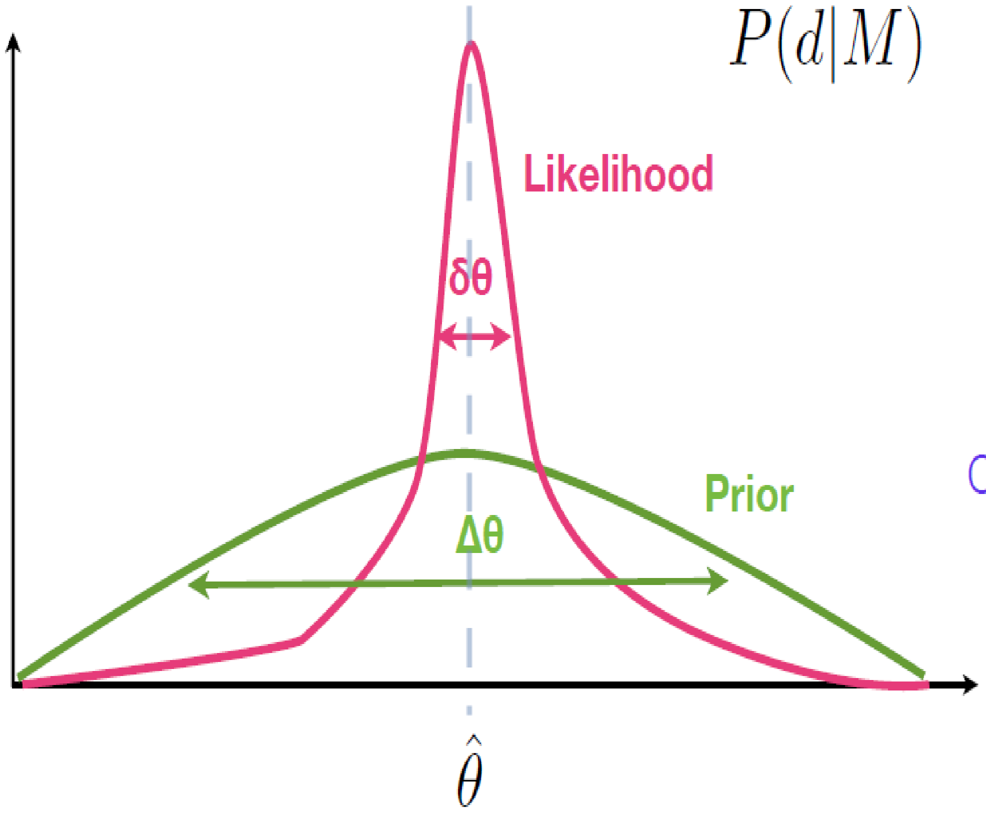
$$p(\vec{d}\,|B) = \int p(\vec{d}|A,B)p(A)dA$$





Integrate over B to get the probability of A

The evidence



$$P(d|M) = \int d\theta L(\theta)P(\theta|M)$$
$$\approx P(\hat{\theta})\delta\theta L(\hat{\theta})$$
$$\approx \frac{\delta\theta}{\Delta\theta} L(\hat{\theta})\hat{\theta}$$

Occam's factor

Evidence is the integral of the likelihood function over the prior

Used to evaluate the relative probabilities of different hypotheses

Again, difficult to calculate for large parameter spaces

Evidence comparison

- Require increase in evidence to warrant more complicated model

$\lvert \ln B \rvert$	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

Advantages of Bayesian methods

- Answers a different, but fundamental question
- Incorporate prior information and all uncertainty
- Complete model
- Model selection through evidence
- Downsides:
  - Computation cost (but this is changing, and changing quickly)
  - Think beyond your posterior distributions

Sampling

Total dimensionality ~ hundreds
Weeks of compute time

Difficult problem!

We want to calculate $P(X\,|\,D, M)$

Non-trivial for non-trivial problems

Have to sample from posterior

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 θ_1

Calculate likelihood L_1

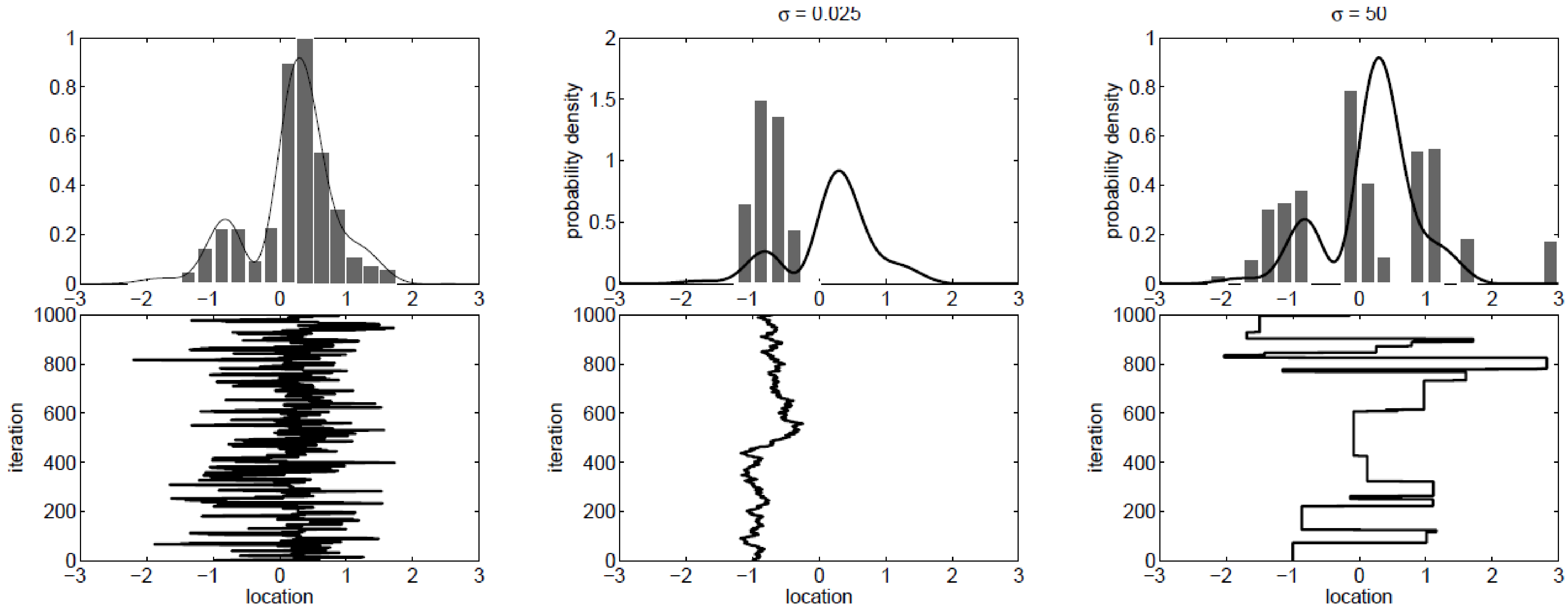
Draw a random uniform number U from 0..1

If L_1/L_0 > U accept the new point, otherwise reject.

Repeat.

Random walk Metropolis Hastings

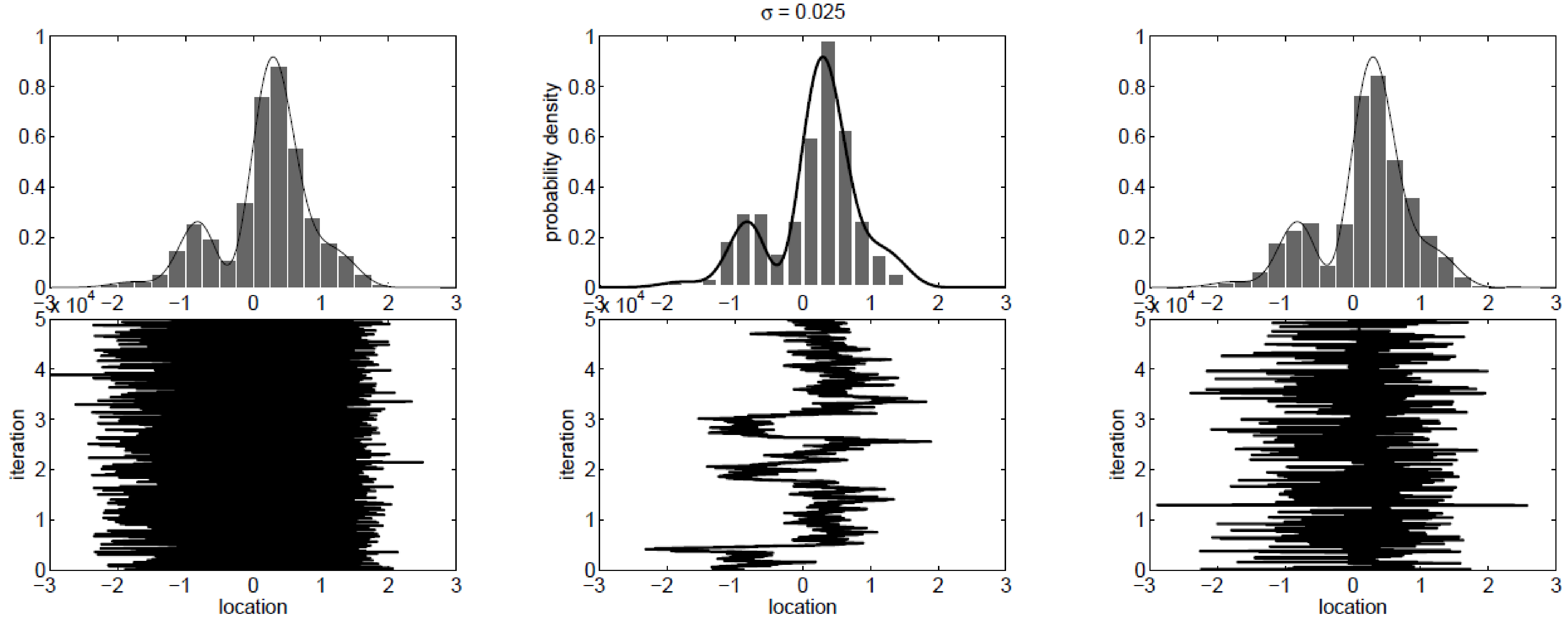
Has its problems: Convergence rate depends on step size



Step Size: Just right Too small Too big

Random walk Metropolis Hastings

But will get there eventually



Step Size: Just right Too small Too large

Random walk Metropolis Hastings

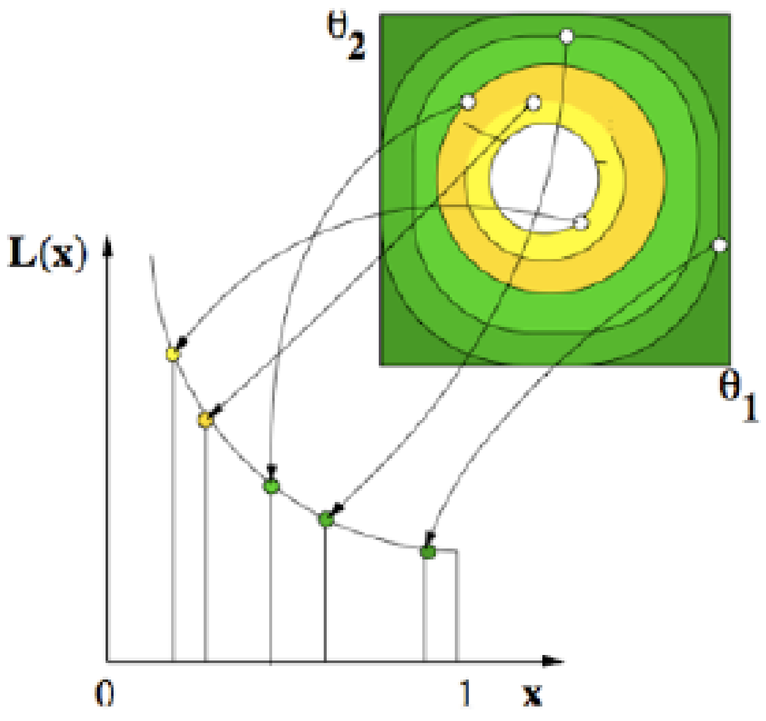
Very poor for multi modal problems:

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

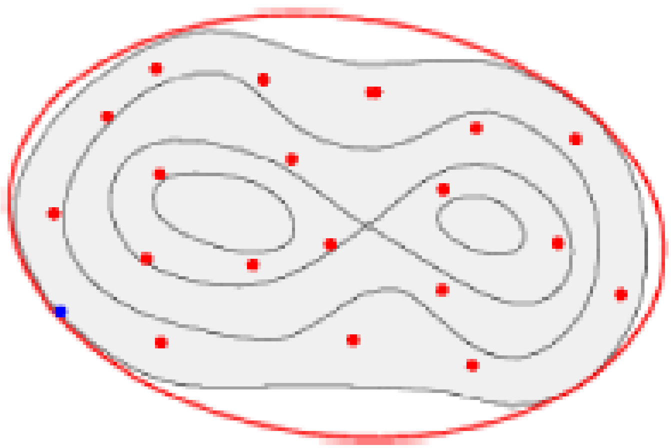
Otherwise try again

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



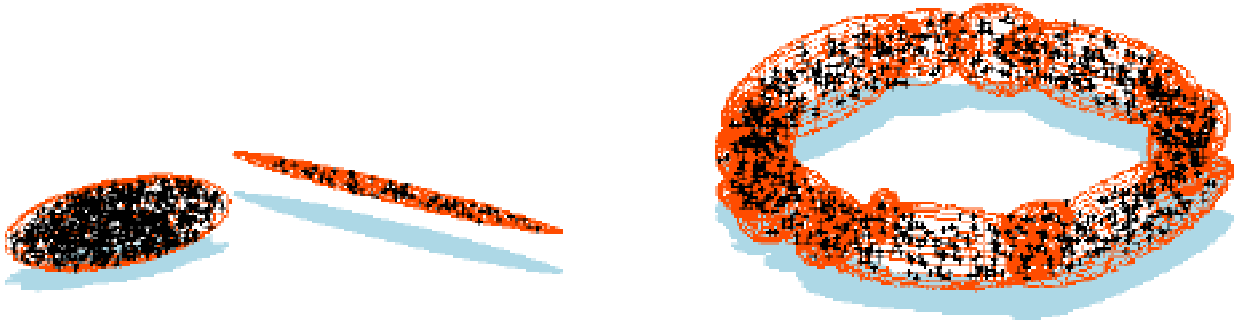
Still not great for multi-modal problems

MultiNest (Feroz & Hobson 2008)

At each iteration:

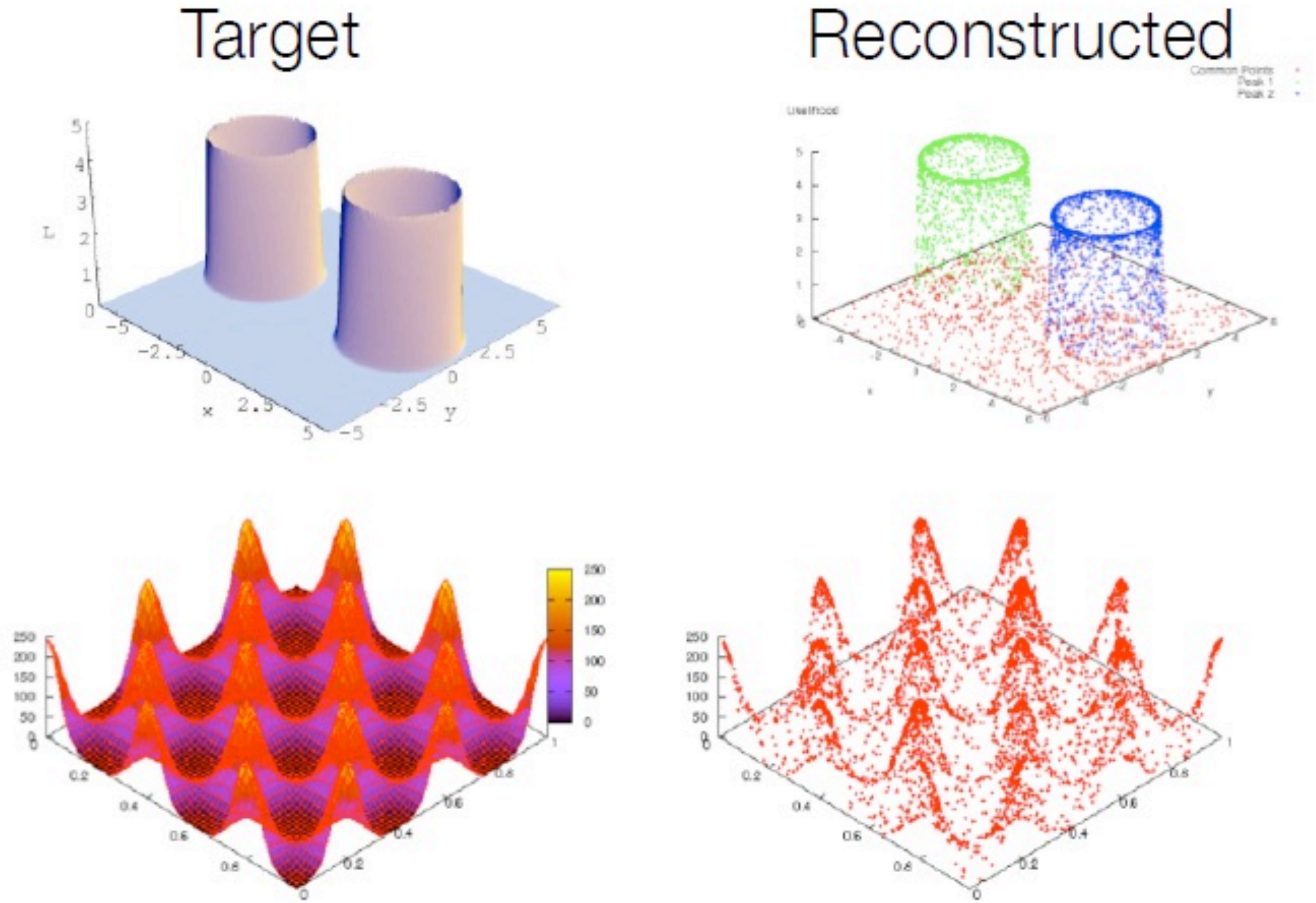
Construct optimal multi-ellipsoidal bound

Pick ellipse at random to sample new point



MultiNest (Feroz & Hobson 2008)

Works great for multi-modal problems:



Polychord (Handley & Hobson 2015)

Successor to MultiNest.

Still uses nested sampling.

Works in much higher dimensions (up to ~ 150)

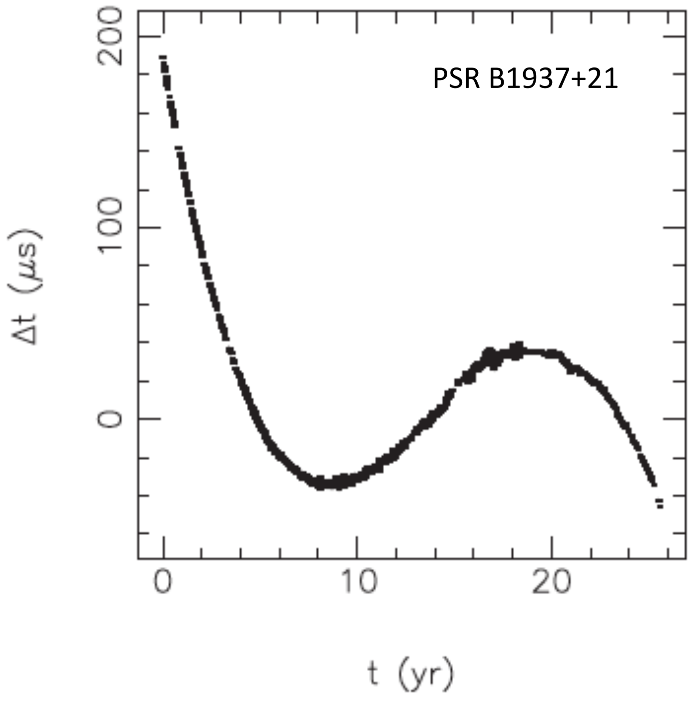
Tests for Gaussianity

$$p(d|\lambda) = \frac{1}{\sqrt{2\pi\sigma^2}} \prod_i \exp\left(-\frac{(d_i - s_i)^2}{2\sigma^2}\right).$$

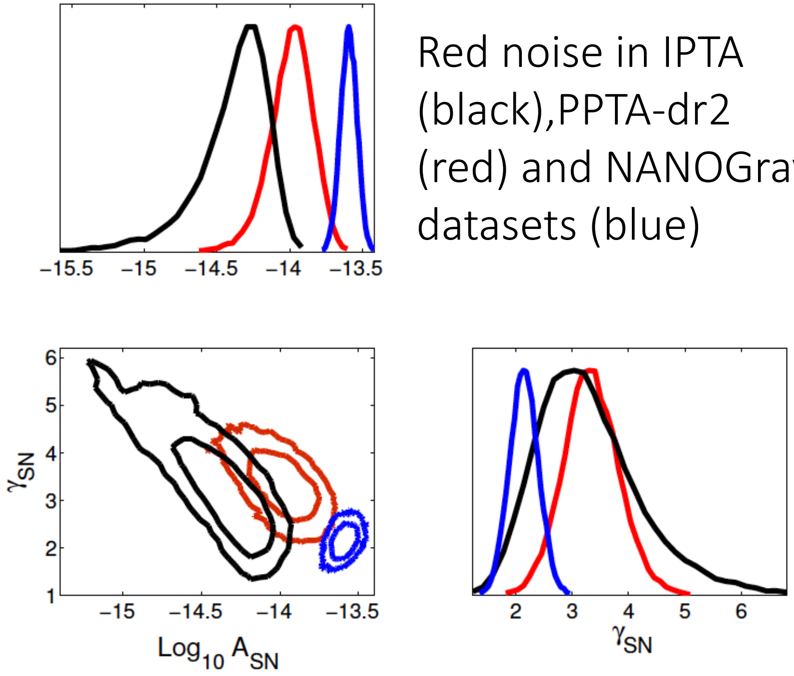
- Why bother?
  - Departures would signify that there is a breakdown of the assumptions about contributions to the measurements
  - Will bias your parameter estimation/posterior distribution
  - Probably indicates contamination from:
    - radio-frequency interference
    - interstellar focusing effects
    - events within the pulsar
- Can do moment tests:
    - asymmetry of the PDF of timing residuals
    - kurtosis, etc.
    - Lentati et al. (2015)

Example: characterizing red noise

Three data sets:
IPTA – first data release
NANOGrav 9 yr
PPTA – same 9 yr as Nanograv



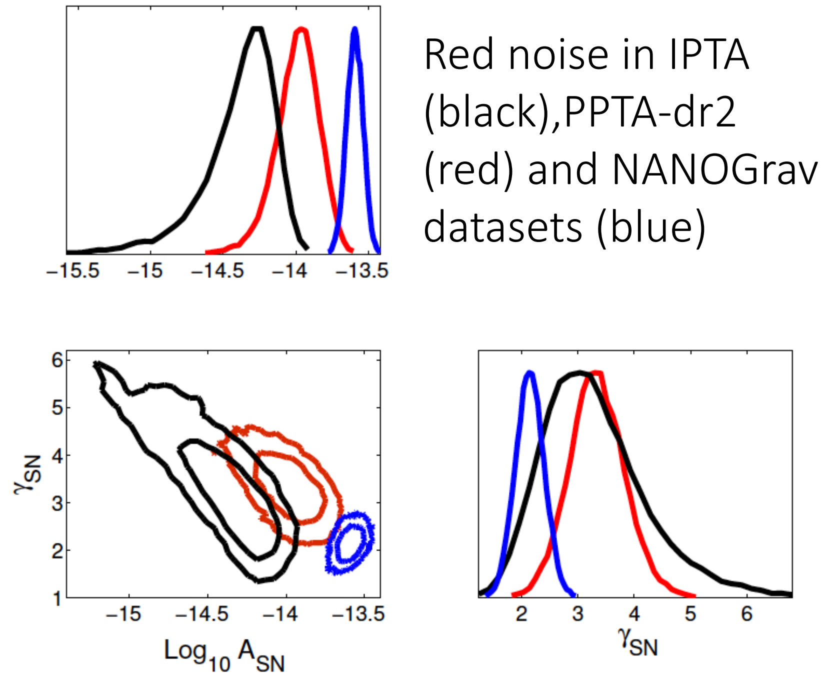
"Triangle plots"



Red noise in IPTA (black),PPTA-dr2 (red) and NANOGrav datasets (blue)

Interpreting results of timing model

- Visual inspection?
  - Do the residuals look weird (sinusoidal trends vs time or orbital phase)
- How good is the fit? How good is the model?
- Are parameters significant? Do their values make sense?
- Do parameters improve the fit/increase the evidence?
- Are parameters consistent?
  - Geometric parallax / vs Change in orbital period
  - Consistent with other observatories, **IPTA**



Red noise in IPTA (black),PPTA-dr2 (red) and NANOGrav datasets (blue)

Ryan’s recipe for precision timing

- Use `tempo2` graphical plugin to inspect residuals
  - “Bailes Method” -> remove the low S/N TOAs and see what is left
  - Sort by frequency/flag by backend, etc.
  - Average data together to see what low S/N signals exist
- Use maximum-likelihood methods to explore data
  - What are the important parameters?
  - What are the important noise sources?
- Use Bayesian methods to explore the models
  - How are parameters covariant?
  - Which parameters/models are supported?

Main stochastic processes in timing models

- White noise
  - EQUAD (TNEQ): Increase all TOA uncertainties by constant value (for example 100 ns)
  - EFAC (TNEF): Adjust TOA uncertainties by a factor
  - ECORR: Correlated error between simultaneously measured TOAs

- Achromatic red noise
  - Power law noise (Amplitude and spectral index)
  - Ephemeris noise (Sarah's talk)
  - Gravitational wave background (Tomorrow)

- Chromatic red noise
  - Dispersion measure variations:
    - Power law process
    - DM events
  - Scattering terms