Updates on Bayesian PTA analysis & the EPTA data analysis pipeline

Rutger van Haasteren
AEI Hannover

June 26th 2012

Outline

General comments on analyses

- Detection & ROC curves
- Parameter estimation & method tests

EPTA data analysis pipeline

Results mock data challenge

Detection and limiting

- Detection: decision rule + diagnostics

  Decision rule: yes / no there was a gravitational wave signal

  Diagnostics: produce Hellings & Downs curve etc.
- Parameter estimation: what are the signal parameters?

  Produce estimators or posteriors
- Limiting = parameter estimation: assuming there is a GW signal in the data, what is the maximum amplitude consistent with the data?

- Bayesian, frequentist, ad-hoc
- False alarm, detection probability...

Unresolvable differences?

Uniform most powerful test (UMP)

Detection methods: answer the question “was there a signal yes/no”

Make an Receiver Operating Characteristic (ROC) curve: false positive vs detection probability

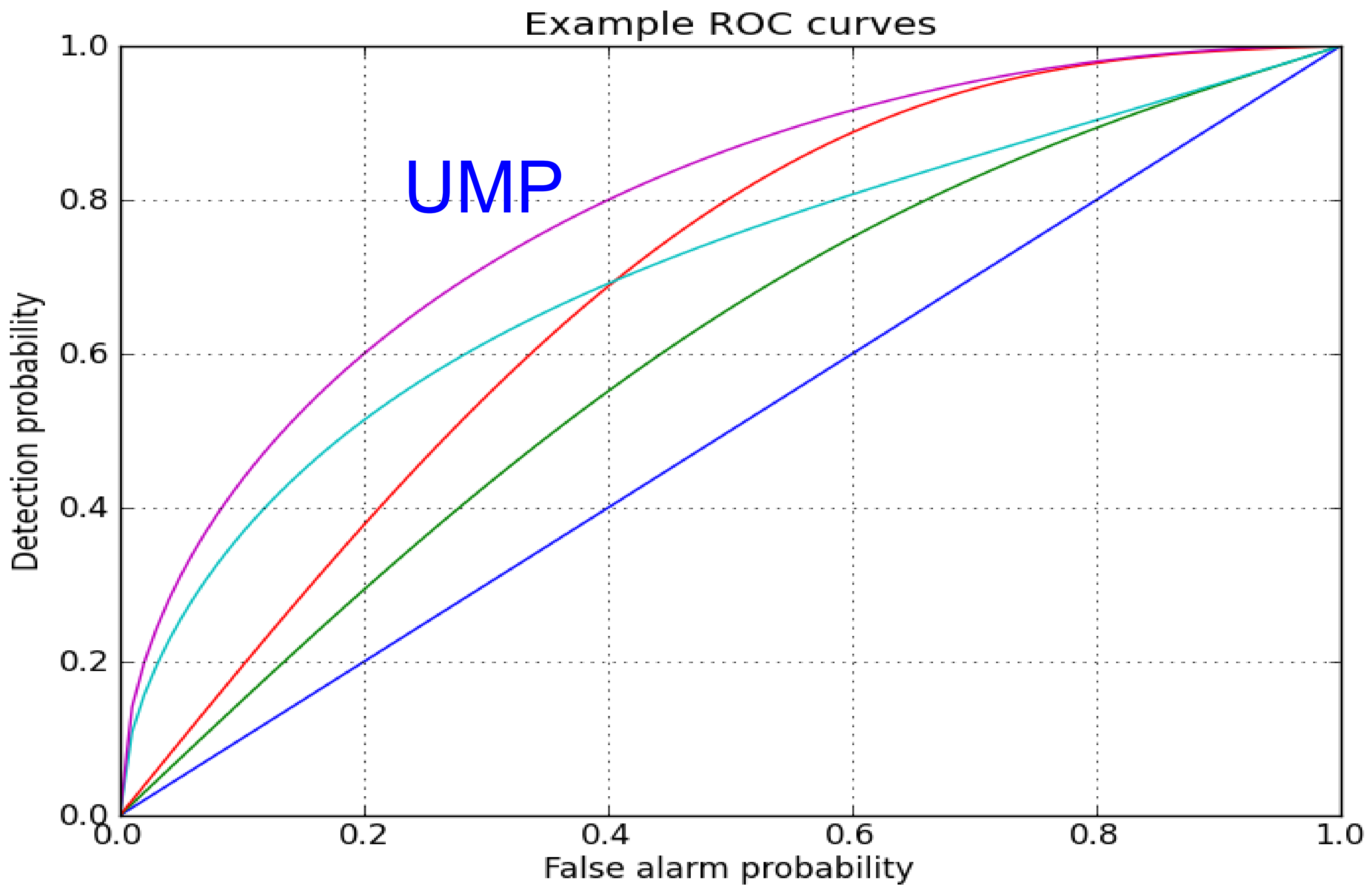
On an ensemble of mock data, with and without a signal, run the detection statistic and calculate the detection probability & false alarm rate.

Uniform most powerful test (UMP) is the test that has the highest detection probability for all false alarm rates.

The UMP depends on the mock data: if other noise characteristics are used the UMP will have a different form.

For simple parametrised point hypotheses, the Neyman-Pearson lemma states that the likelihood ratio is the UMP. But is it possible to derive the UMP in general??

ROC Curve



Example ROC curves
UMP
Detection probability
False alarm probability
0.0
0.2
0.4
0.6
0.8
1.0

Uniform most powerful test (UMP)

But is it possible to derive the UMP in general??

- Yes!!! That quantity is called the 'Bayes factor'.
- The priors used in calculating the Bayes factor are the distributions used in realising the ensemble of mock datasets.
- Different priors lead to:
  - A different Bayes factor for Bayesians
  - Different mock datasets and therefore a different UMP for Frequentists.

Unresolvable differences?

Bayes factor is the UMP statistic if the same priors are used for frequentist and Bayesian methods

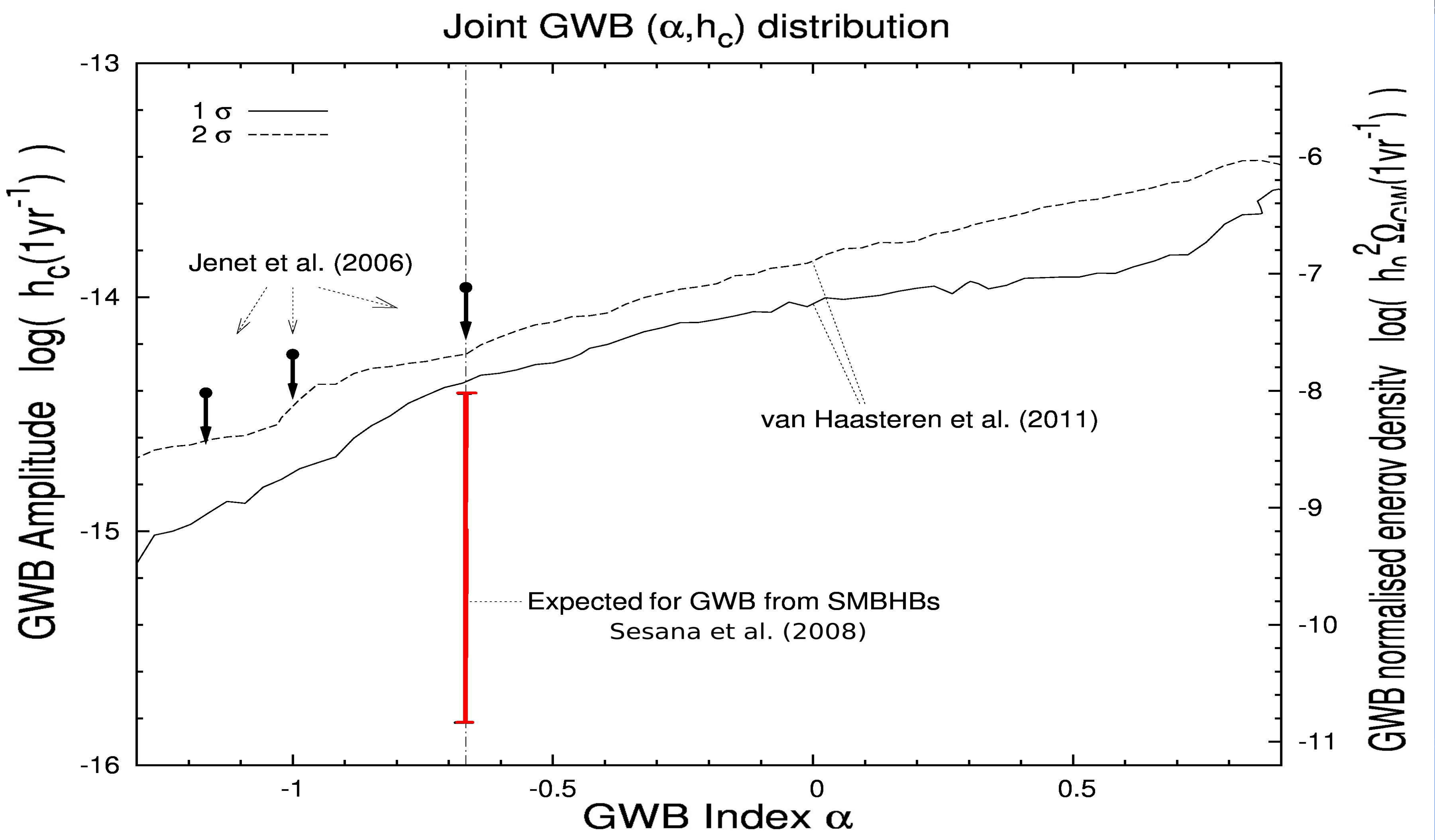
The big question of detection

What simulations do we use to generate our ROC curves?

Should include all effects that could possibly mimick a GW(B) that we want to be insensitive to (e.g. Clock errors).

Our models are by far not accurate enough to derive a good detection statistic. And even if they were, our current detection statistics are either suboptimal, or too computationally expensive to calculate.

Parameter estimation

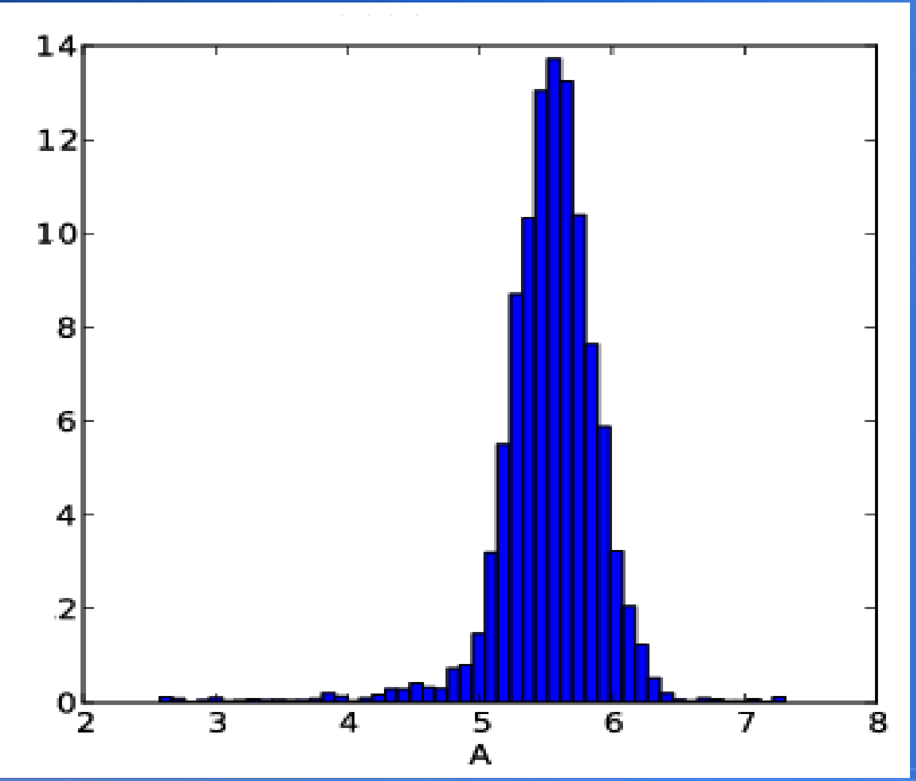


Parameter estimation

How well can we estimate the value of a parameter, and can we trust the uncertainties? Simulations!

- Frequentists look for accurate estimators: compare estimator with true value for lots of simulations

- Bayesians look for accurate posteriors: how do we test correctly with simulations??



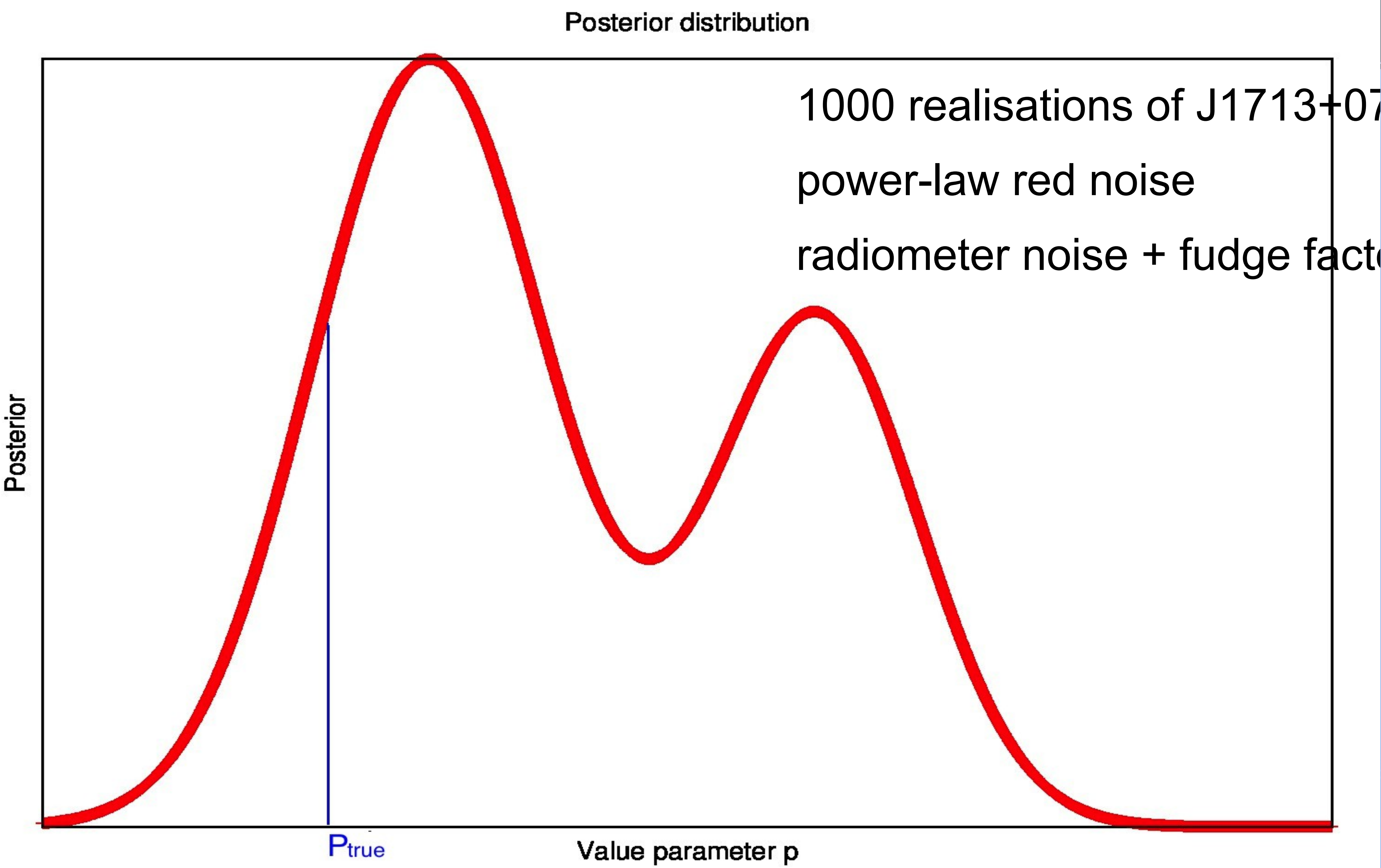
Example: red noise analysis

Method from van Haasteren & Levin, submitted. Similar to Cholesky method: a red noise model used to obtain solution for timing model.

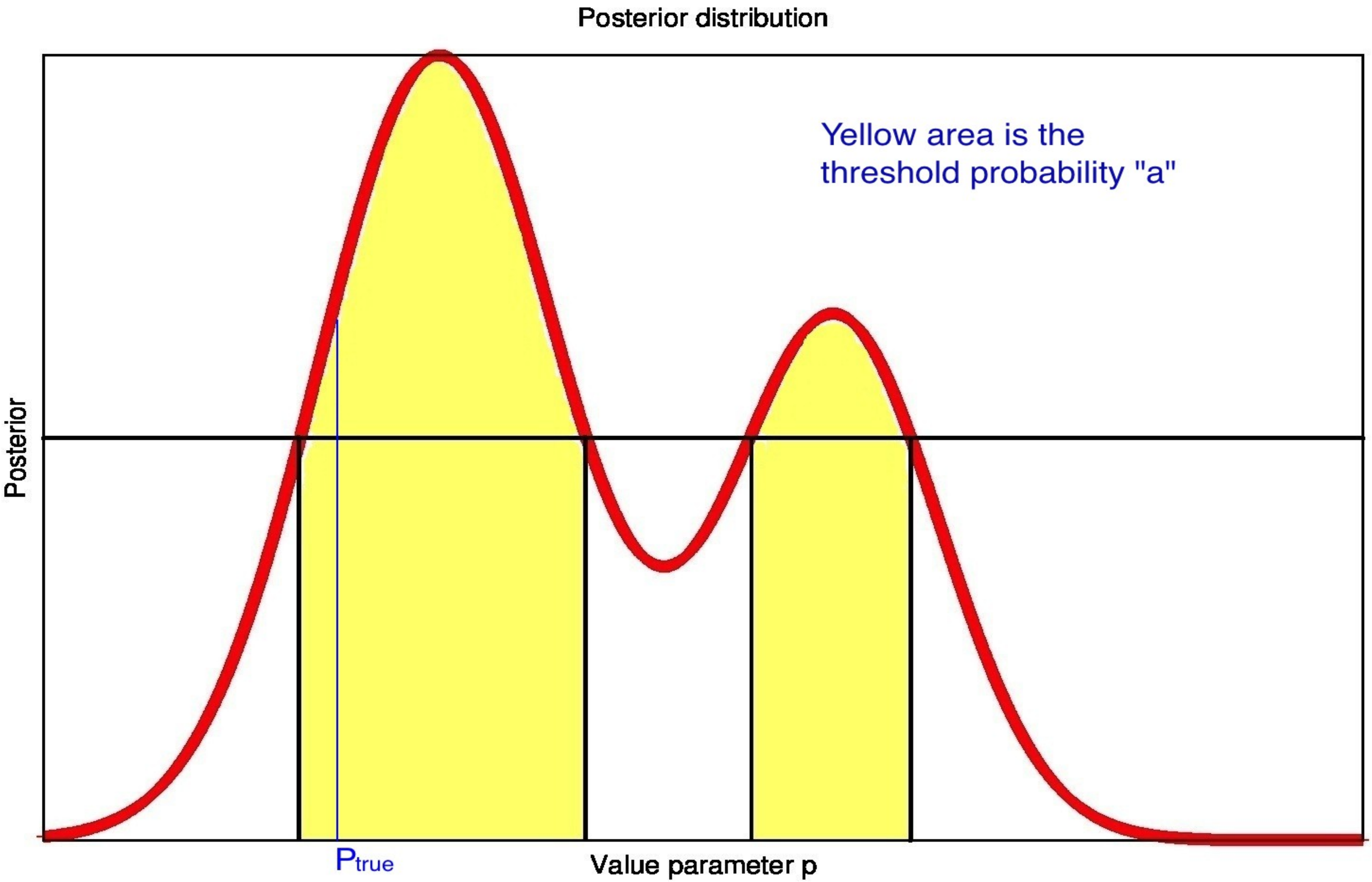
- No computational overhead
- Accurate estimates for both red noise & timing model
- No approximations. Statistically correct if model is correct

Coles et al. (2011) compared their estimates for n=100 realisations of mock data. Rms of estimates was in agreement with true values for all timing model parameters, except for F0 and F1.

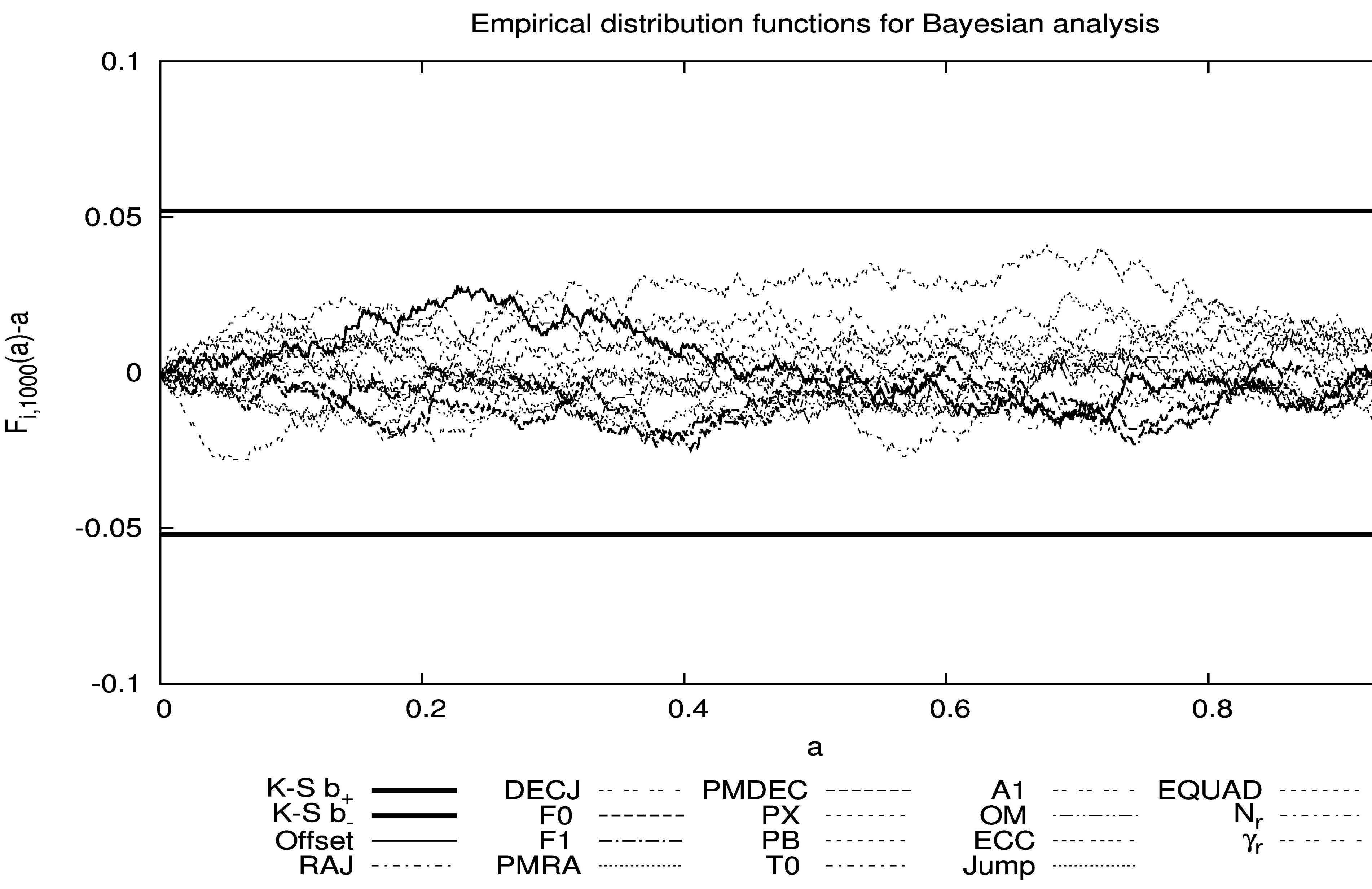
Posterior distribution



Threshold probability a

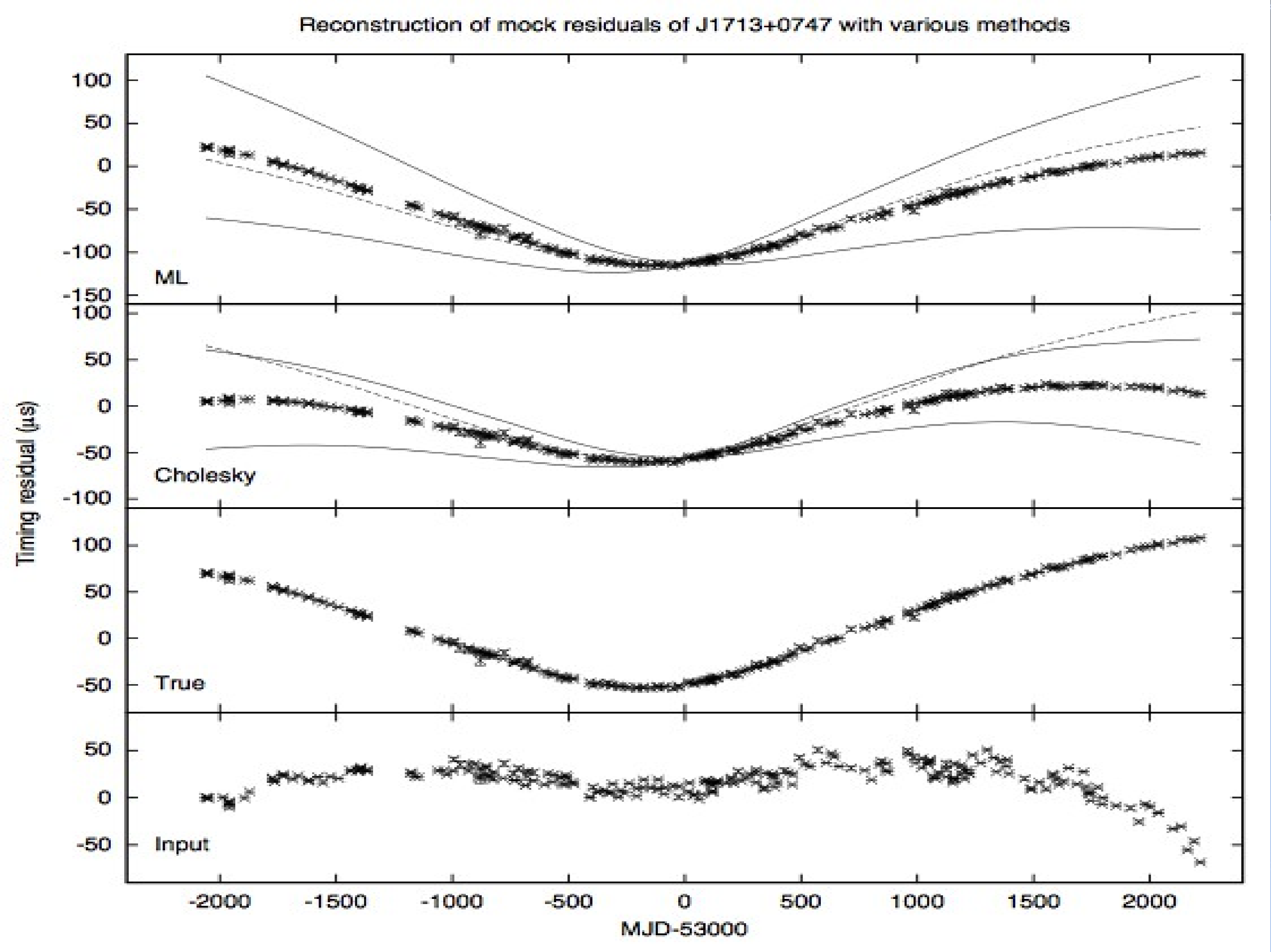


Kolmogorov-Smirnov test



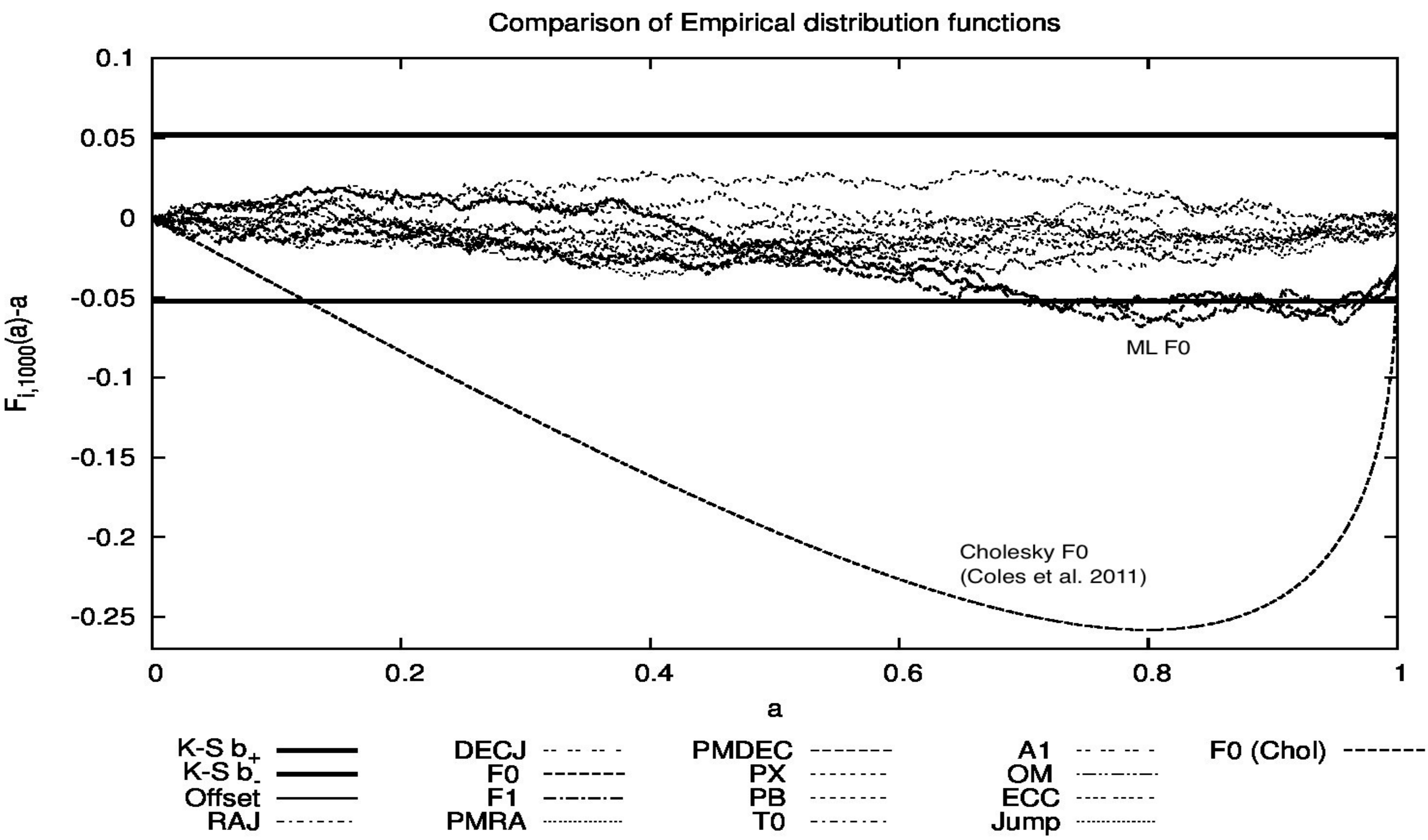
Extra notes

- In principle, we *knew* this would be the answer. There were no approximations! But, good test for implementation
- Works for all model parameters
  - Binary parameters
  - Frequency & frequency derivative
  - Red noise parameters
- Of course, whether our model is good is another thing
- Sensitive test for consistency of analysis method.
- Works for frequentist estimators, too! Example...



Reconstruction of mock residuals of J1713+0747 with various methods
ML
Cholesky
True
Input
Timing residual (μs)
MJD-53000
100
50
0
-50
-100
-150
-2000
-1500
-1000
-500
0
500
1000
1500
2000

Getting the covariance right



GWB is very low-frequency!

- Lowest frequencies are most important for optimal GW(B) detection
- Covariance matrix is essential to get right. Also: non-stationary!
- Timing noise non-stationary on longer time scales (e.g. DM correction done yes/no)
- Clock errors

Bayesian/EPTA pipeline

We want to search for different kinds of waves...

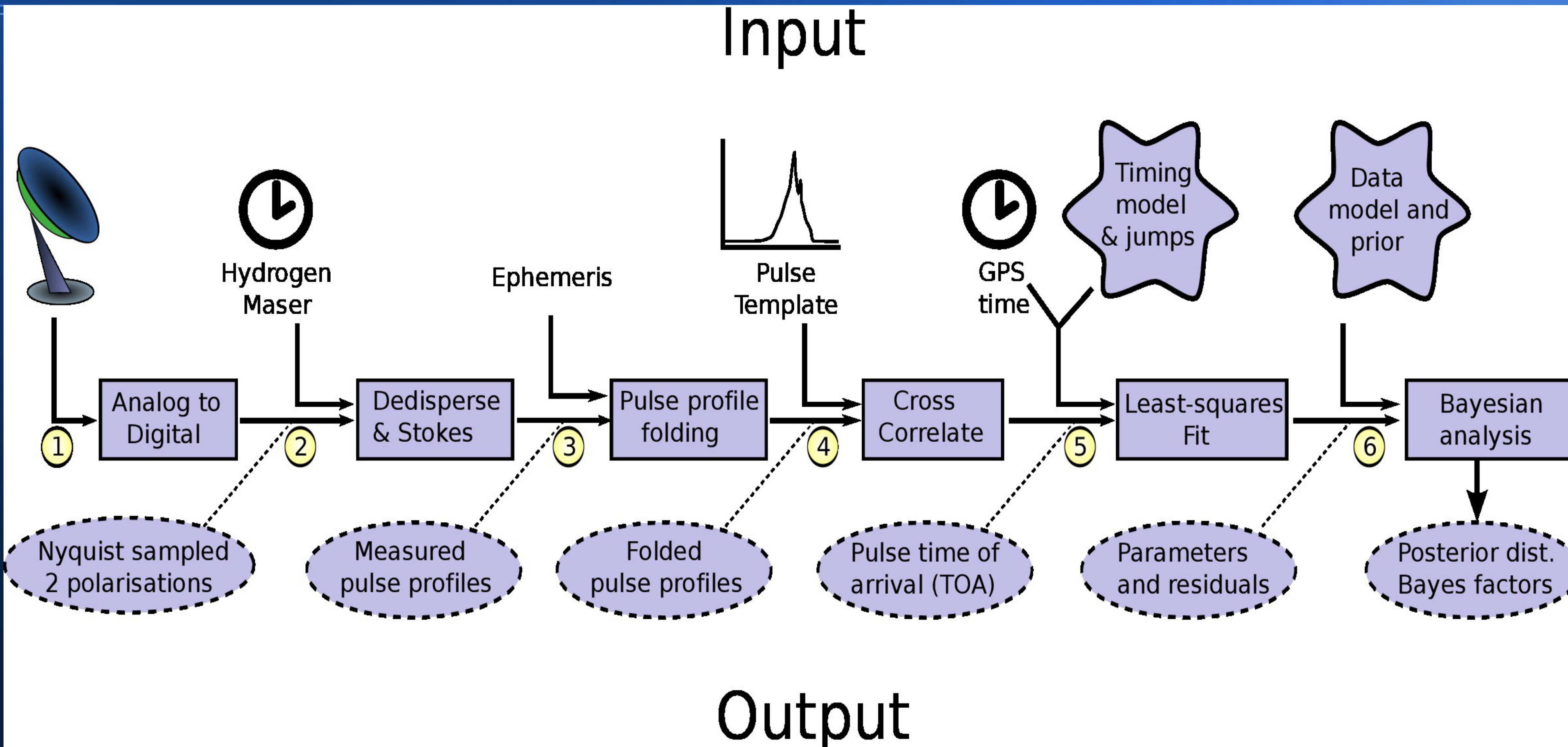
Under construction

What do we have?

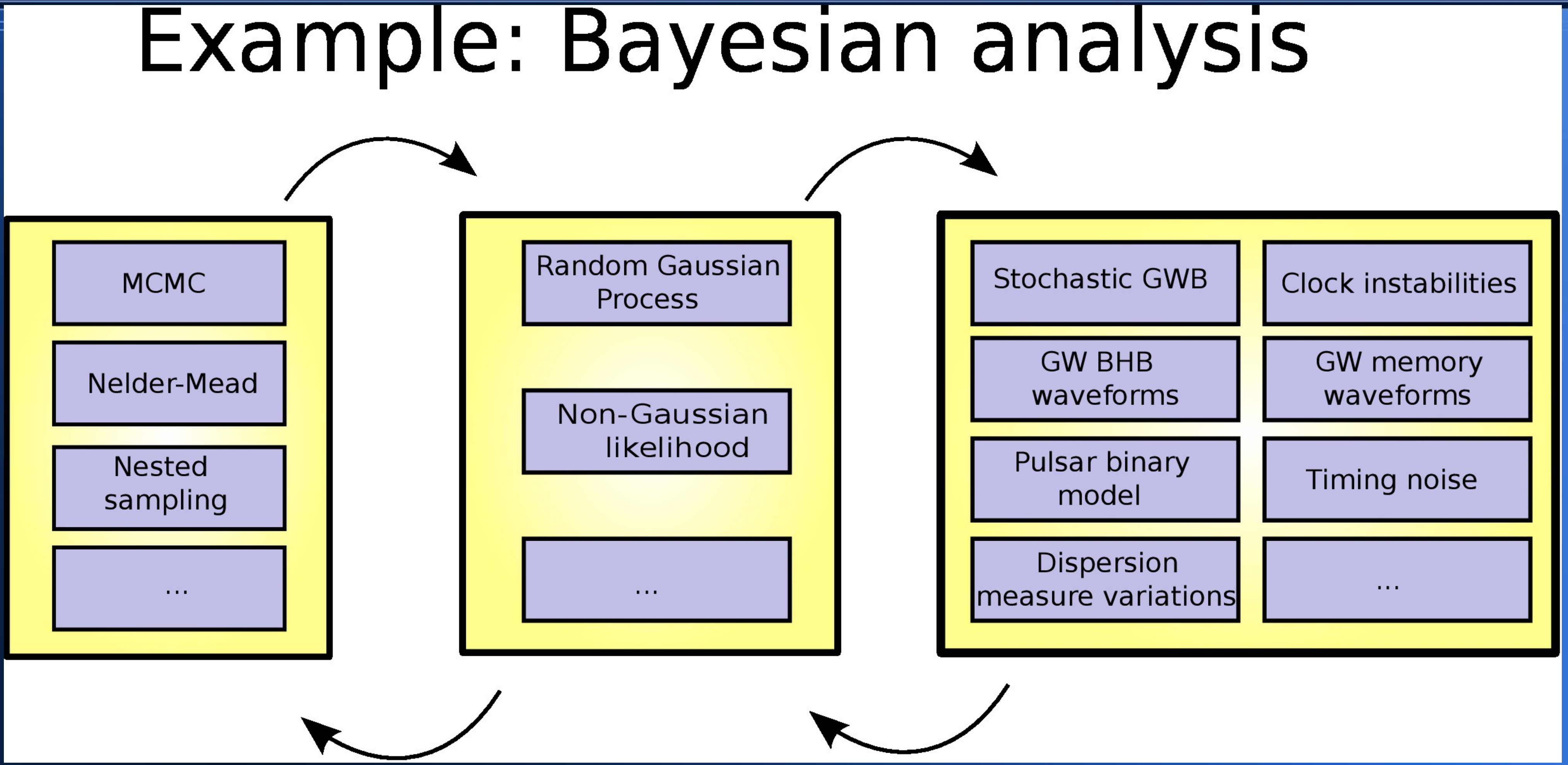
Bayesian/EPTA pipeline

- Implemented/tested in some form:
  - Random Gaussian process
  - Noise: white (radiometer, jitter, …) red noise, clock errors
  - Linearised timing moded (from Tempo2)
  - Isotropic GW background
  - GW memory
  - Non-evolving BH binaries
  - Optimal DM correction (see poster K.J.)
  - ...
- Not all part of same pipeline yet
  - Include other work (Babak, Sesana, Mingarelli, Vecchio, Lassus, ...)
  - Non-Gaussian noise
  - ...
- We need one environment to do timing data analysis without having to re-do work all the time! EPTA data analysis project: see poster by Antoine

EPTA data analysis library: 5&6



EPTA data analysis library



EPTA data analysis project

Antoine Lassus, K.J. Lee, Chiara Mingarelli, Alberto Vecchio, Rutger van Haasteren

Open source as soon as possible

Written in Python/C

Should become a library for all kinds of timing data analysis, aimed both at rapid implementation of new methods, and routine work

Mock data challenge

One telescope: virtual axis telescope

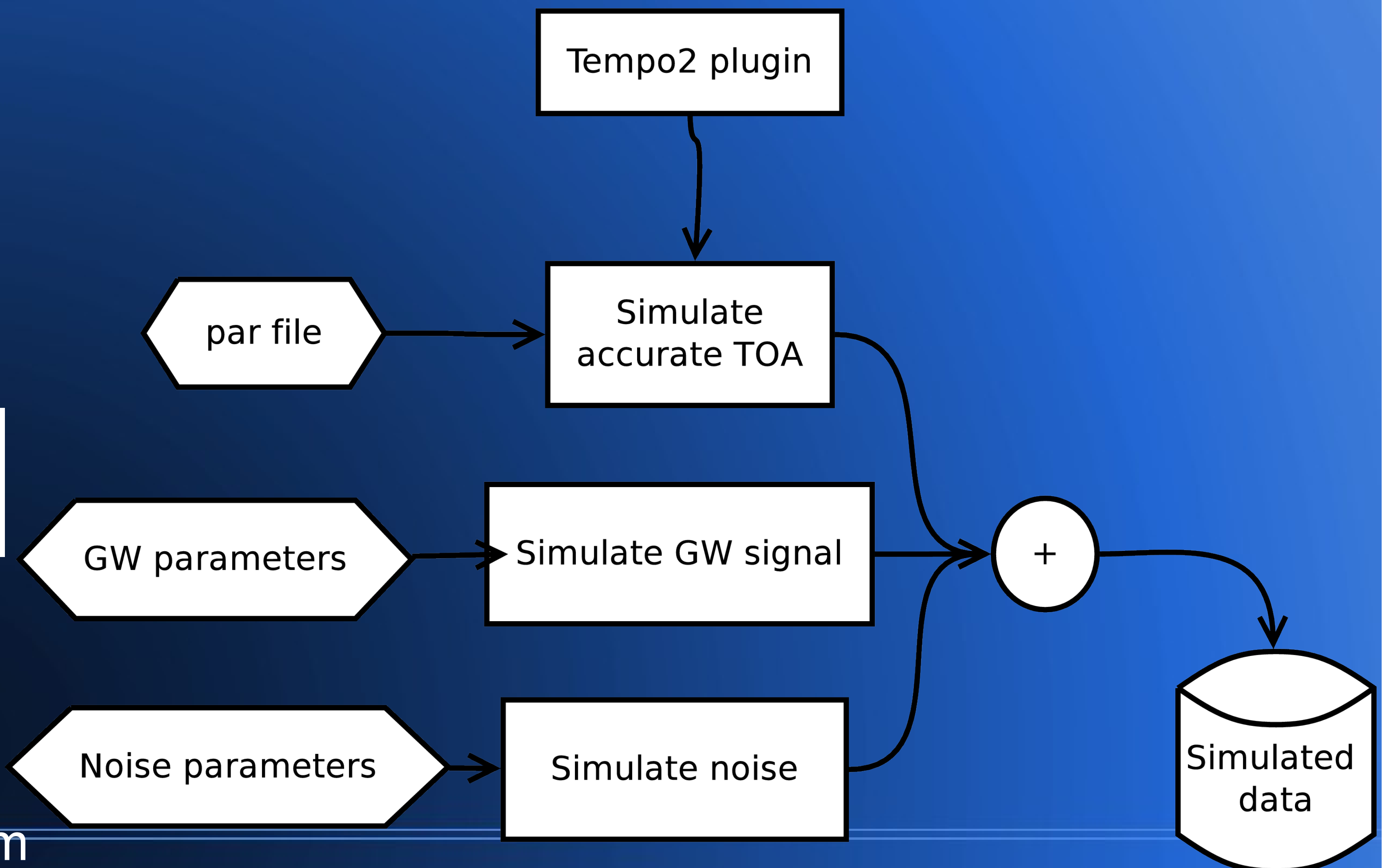
36 pulsars, overlapping dataset, no gaps

Timing noise simple: power-law + radiometer (error bars)

No jumps: only one receiver used

Isotropic GWB with spectral index -2/3

Conversions, code implementation and algorithm



$$h_c(f) = A_c \left(\frac{f}{1\mathrm{yr}} \right)^{-\alpha}$$

GW spectrum

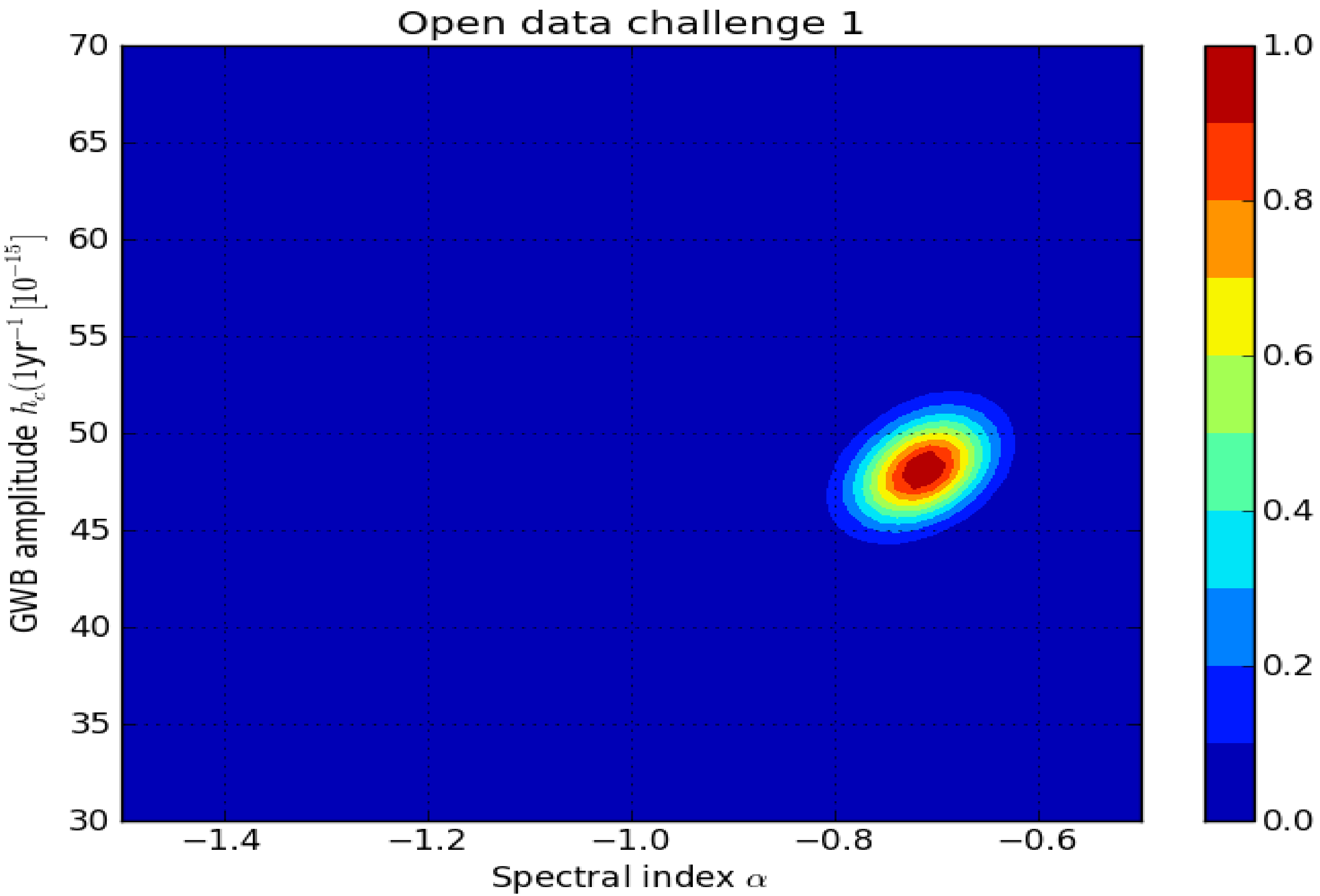
$$P(f) = A f^{-\alpha}$$

Noise spectrum

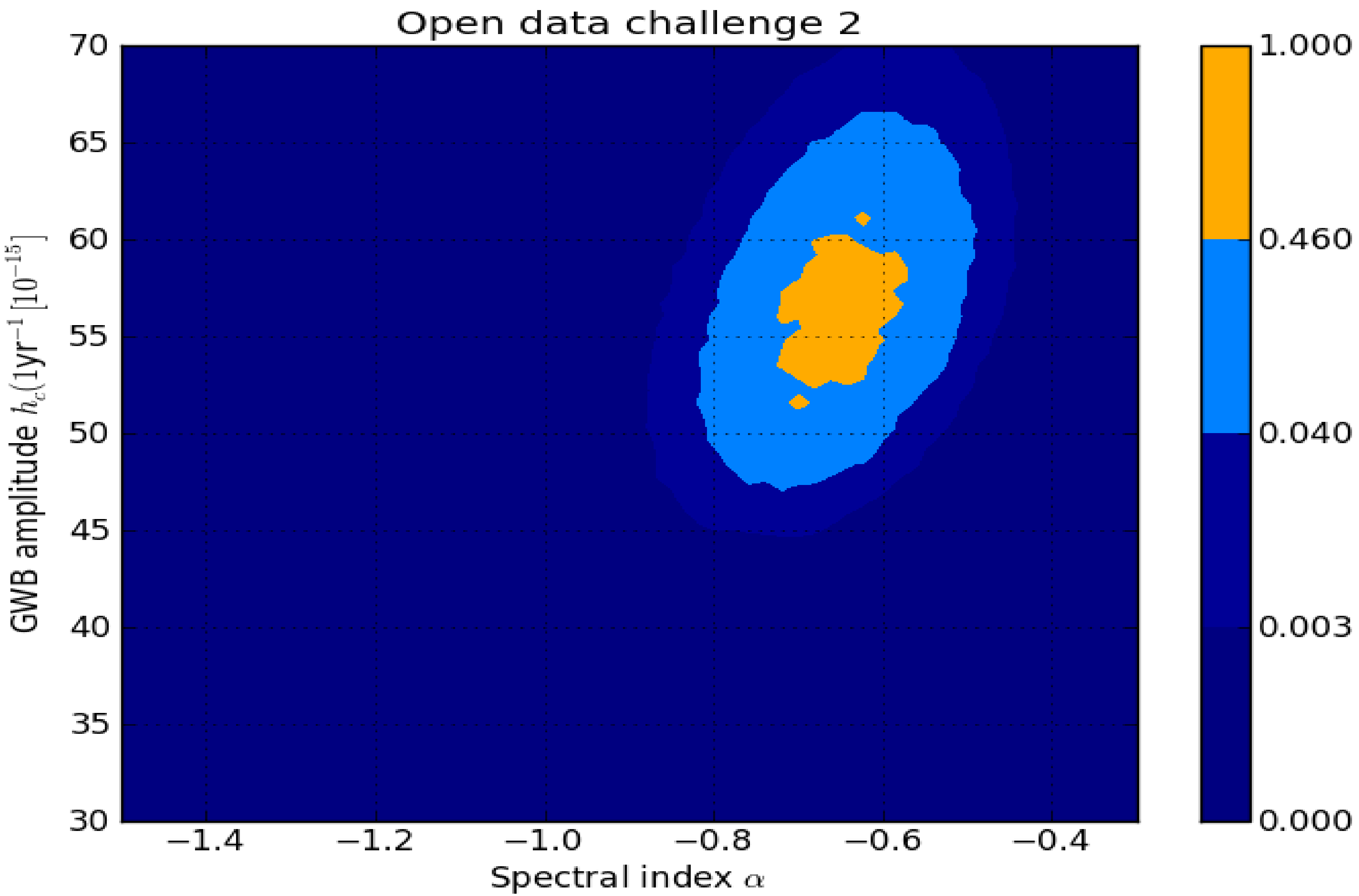
Results

- Given the challenge, we will treat it as a parameter estimation problem: either limit or characterise the GWB signal in the data
- Model:
  - Random Gaussian process
  - TOA error bars (no fudge factors)
  - Tempo2 timing model
  - Power-law red noise
  - GWB with H&D correlations and power-law PSD

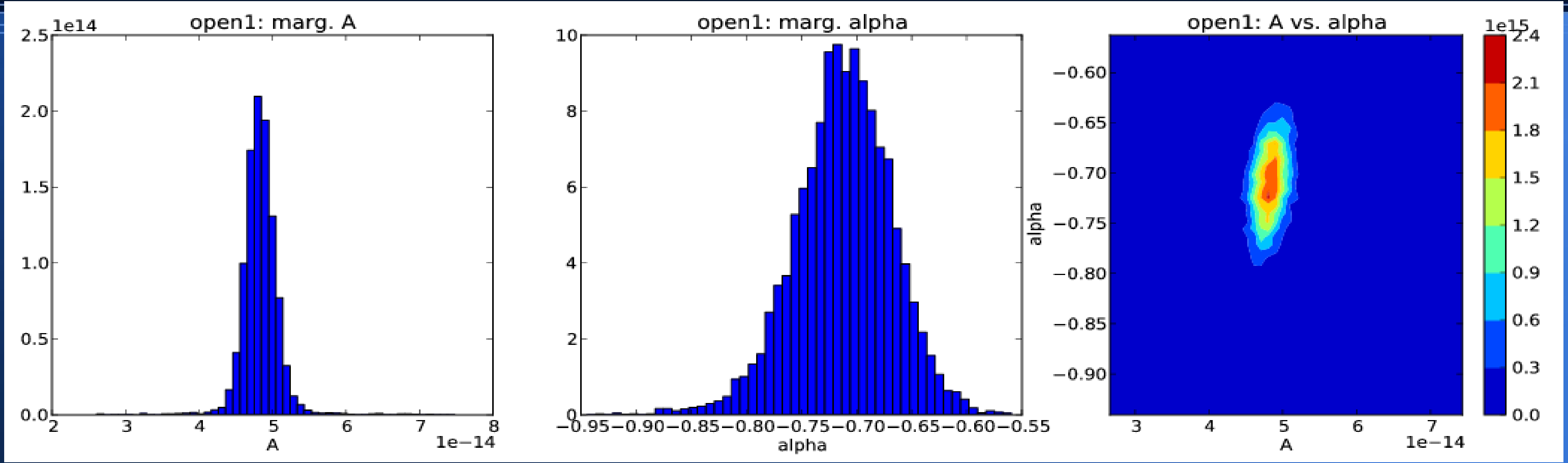
Results open challenge

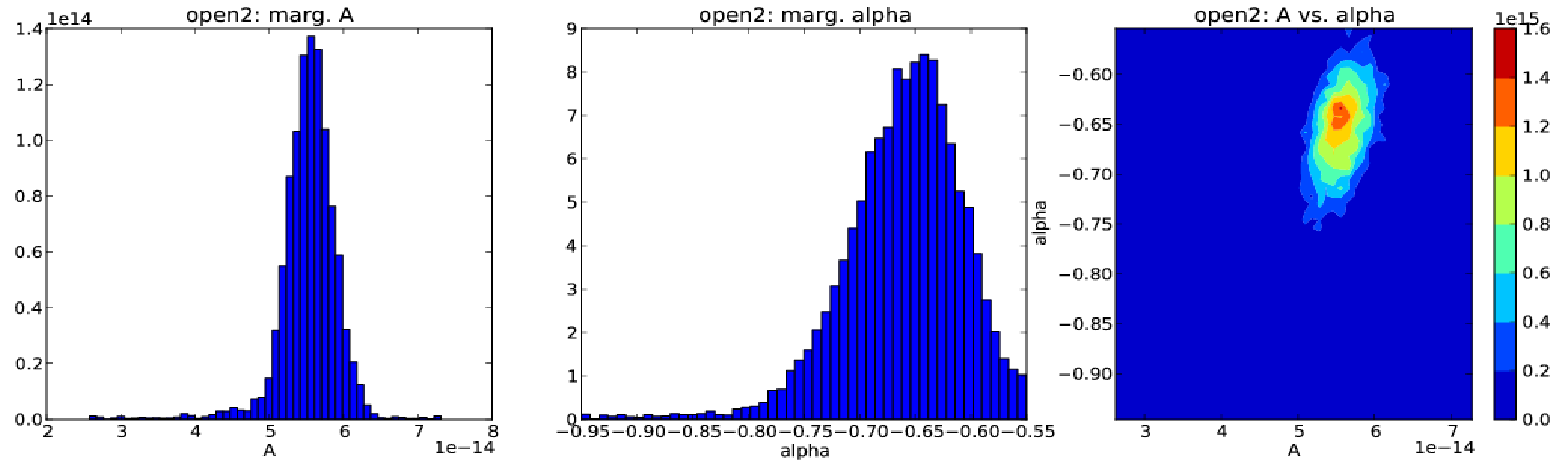


Results open challenge



Independent implementation by Michele Vallisneri (JPL)





Conclusions

Detection: Bayesian and Frequentist UMP test the same. Produce ROC curves!

- We should start comparing those as well.
- Data challenge very good first start!

Parameter estimation: K-S test

- Are methods accurate enough
- Are lowest frequencies handled correctly?

EPTA data analysis pipeline (see poster by Antoine). Implemented in Python/C

Data challenge approach as a parameter estimation problem. Different Bayesian analysis implementations already been tested, and results of open set have been confirmed to be correct.