Model selection with Markov-Chain Monte-Carlo results applied to Gravitational Waves

V. Raymond[1], M.V. van der Sluys[2], I. Mandel[1], V. Kalogera[1], C. Röver[3], N. Christensen[4]

[1]Northwestern University, [2]University of Alberta, [3]MPI Hannover, [4]Carleton College. E-mail: vivien@u.northwestern.edu

1 Introduction

Among the possible sources of gravitational waves for ground based interferometers, binary compact object inspirals with members in the mass range $1M\odot - 100M\odot$ are likely detection candidates and relatively easy to model. A gravitational-wave event detection is challenging, and will be full of astrophysically relevant informations. Bayesian analysis of source properties holds major promise for improving our astrophysical understanding and requires reliable methods for parameter estimation and model selection. Markov-Chain Monte Carlo (MCMC) is used to obtain the probability-density function (PDF) of the parameter vector. The MCMC output is also used to compute the Bayes factor of spinning and non-spinning models fitting data sets in which a spinning signal was injected.

2 The waveform

Our waveform model computes templates up to 3.5 pN order in phase (Newtonian in amplitude) [1], using the LSC Algorithm Library.

$$\vec{\lambda} = \{\mathcal{M}_c, \eta, d_L, t_c, \varphi_c, \alpha, \delta, \iota, \psi, a_{spin1}, \theta_1, \phi_1, a_{spin2}, \theta_2, \phi_2\}$$

The 15 parameters: chirp mass, mass ratio, distance, time and phase at coalescence, right ascension and declination, inclination and polarisation angle, magnitude, inclination and phase of spin 1 and 2.

3 Methods

Given a model M depending on parameter set $\vec{\lambda}$, the prior on this set $p(\vec{\lambda}|M)$ and the data $\vec{x}$, the posterior PDF of the set is given by:

$$p(\vec{\lambda}|\vec{x}, M) = \frac{p(\vec{\lambda}|M)\, p(\vec{x}|\vec{\lambda}, M)}{p(\vec{x}|M)} \qquad \left(posterior = \frac{prior * likelihood}{evidence}\right)$$

We use MCMC [2,3] to explore the parameter space. The histogram of the chain converges towards a function proportional to the PDF.

4 Parallel Tempering

We run several chains in parallel and artificially modify the likelihood function with the "temperature" T:

$$p_T(\vec{x}|\vec{\lambda}, M) = p(\vec{x}|\vec{\lambda}, M)^{T^{-1}}$$

Additional proposals allow swapping between chains [5]. In fig.1a the posterior value for different temperature chains as a function of the iteration number is shown, with the corresponding PDF in fig.1b. A simple multi-dimensional Gaussian likelihood is used.
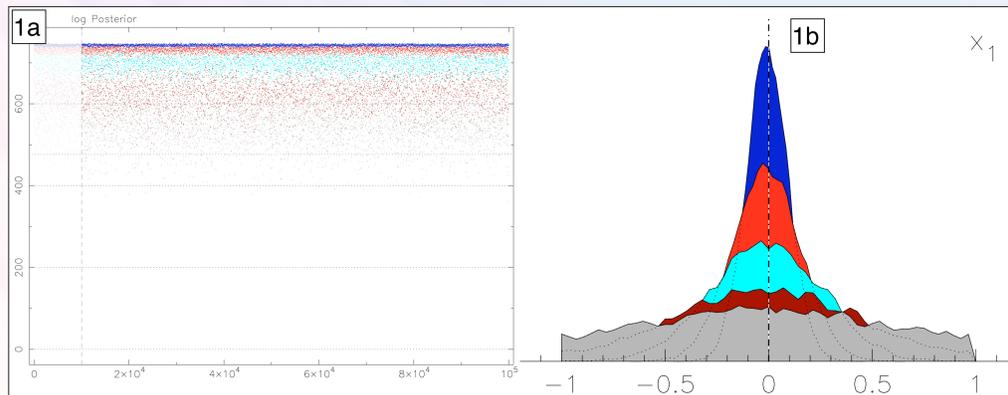

Fig.1. **MCMC posterior values for a gaussian likelihood** (1a) and corresponding marginalized PDF, with only one parameter shown (1b). The colors correspond to different temperatures. From cold to hot: **T=1.0** , **T=2.7** , **T=7.1** , **T=18.8** , **T=50.0** .

5 Evidence

Bayes' theorem on the Mode M_i: $\quad p(M_i|\vec{x}) = \dfrac{p(M_i)\, p(\vec{x}|M_i)}{p(\vec{x})}$

allows us to use the data available to compare two models M_i and M_j:

$$O_{i,j} = \frac{p(M_i|\vec{x})}{p(M_j|\vec{x})} = \frac{p(M_i)\, p(\vec{x}|M_i)}{p(M_j)\, p(\vec{x}|M_j)} = \frac{p(M_i)}{p(M_j)} B_{i,j}$$

We need to compute the *evidence* $p(\vec{x}|M_i)$, which for a parametrized model is an integral over the parameters. This integral is approximated by a discrete sum over the sampled points [4]:

$$p_T(\vec{x}|M_i) \approx \sum_{k=1}^{N} p(\vec{\lambda}_k|M_i)\, p(\vec{x}|\vec{\lambda}_k, M_i)\, V_{T\vec{\lambda}_k}$$

Where $V_{T\vec{\lambda}_k}$ is the volume associated with point $\vec{\lambda}_k$. The sampling properties of the MCMC ensure that:

$$\lim_{N\to\infty} V_{T\vec{\lambda}_k} = \alpha_{Ti} \left(p(\vec{\lambda}_k|M_i)\, p_T(\vec{x}|\vec{\lambda}_k, M_i)\right)^{-1} = \alpha_{Ti} \left(p(\vec{\lambda}_k|M_i) \left(p(\vec{x}|\vec{\lambda}_k, M_i)\right)^{T^{-1}}\right)^{-1}$$

The total parameter space volume $V_{Tt} = \sum_{k=1}^{N} V_{T\vec{\lambda}_k}$ yields a value for α_{Ti} and the evidence is approximated by:

$$p_T(\vec{x}|M_i) \approx V_t \left[\sum_{k=1}^{N} \left(p(\vec{\lambda}_k|M_i)\, p(\vec{x}|\vec{\lambda}_k, M_i)\right)^{-1}\right]^{-1} \sum_{k=1}^{N} p(\vec{x}|\vec{\lambda}_k, M_i)^{1-T^{-1}}$$

Fig.2 shows the evidence for the Gaussian likelihood of section 4 as a function of number of points used (2a) and the spread for different ways of combining the evidences of temperature chains (2b). The geometric mean seems to be the least biased and with smallest variance in this limited test.
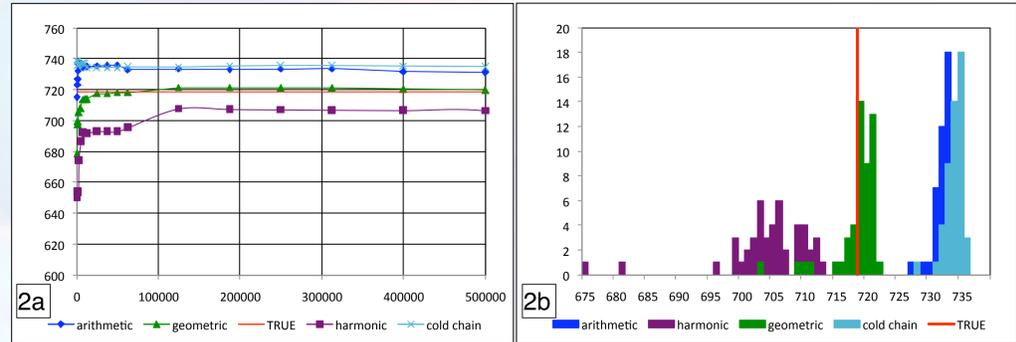

Fig.2. **Evidences** as a function of iteration number (2a) and their spread out of 50 trials (2b). The colors represent different means of the evidences from each temperature chains.

6 Detector data

We present results on LIGO detector data (H1L1) where a spinning signal was injected (M_1=10$M_\odot$, M_2=5$M_\odot$, a_{spin1}=0.8, a_{spin2}=0.5) at SNR=13, and on a coincident artefact (*glitch*) in H1L1 at the same SNR. Fig.3 shows the marginalized PDFs, and the Bayes factors of the models *signal+noise* versus *noise only* are in table 1. As expected the Bayes factor is higher for more accurate models. Correlations among parameters (e.g. sky position) may also help differentiate glitches from signals (fig. 3b).
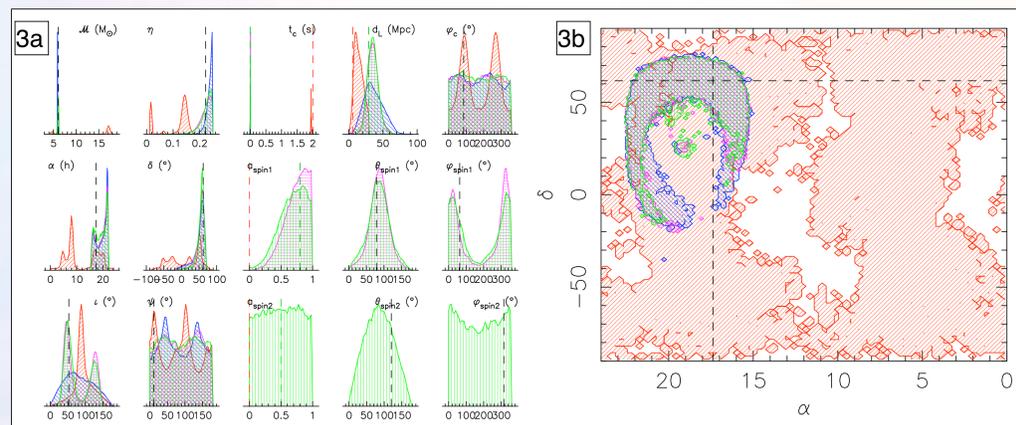

Fig.3. **Marginalized 1D PDFs** (3a) and **2D PDF of the sky position** (3b). The colors correspond to different data sets: **blue** , **purple** , **green** , for an injection at SNR=13 in detector noise H1L1 recovered with **no**, **1** or **2** spins. **Red** , for a **glitch** of SNR~13 in H1L1 recovered with no spins (spinning templates did not converge on the glitch).

templates data sets	no spin		1 spin		2 spins	
synthetic noise	52.4	34.3	68.9	41.8	69.0	43.5
detector noise	60.6	42.3	67.6	45.6	68.6	45.7
glitch	37.7	30.2	74.5*	36.9*	83.2*	37.4*

Table 1. **Logarithms of Bayes factors of signal versus noise**. 3 templates are used on 3 data sets. In each cell the left number was computed using the coldest chain, while the right number using the geometric mean of the evidences from all the chains. The numbers with (*) mark are from chains not having converged, and suffer from under-sampling.

7 Conclusions

• Given our understanding our MCMC allows model selection.

• The geometric mean seems to give the best results. We are investigating this on various posterior shapes and in more details. Further improvements are possible.

• In table 1, the small improvement between the analyses with 1 and 2 spins is due to the small effect of the second member of the binary on the waveform due to its low mass.

• The analyses of the glitch using spinning templates do not converge. This may provide suspicion against a real signal and warrants further studies.

References:
[1] Buonanno A, Chen Y, & Vallisneri M, PRD, 67, 104025 (2003)
[2] Van der Sluys M, Raymond V, Mandel I, Röver C, Christensen N, Kalogera V, Meyer R & Vecchio A, CQG, 25, 184011, (2008)
[3] Van der Sluys M, Röver C, Stroeer A, Raymond V, Mandel I, Christensen N, Kalogera V, Meyer R & Vecchio A, ApJ, 688, L61
[4] Newton M A, Raftery A E, Journal of the Royal Statistical Society, Series B, 56, 3 (1994)
[5] Earl D J, Deem M W, Physical Chemistry Chemical Physics, 7, 3910 (2005)