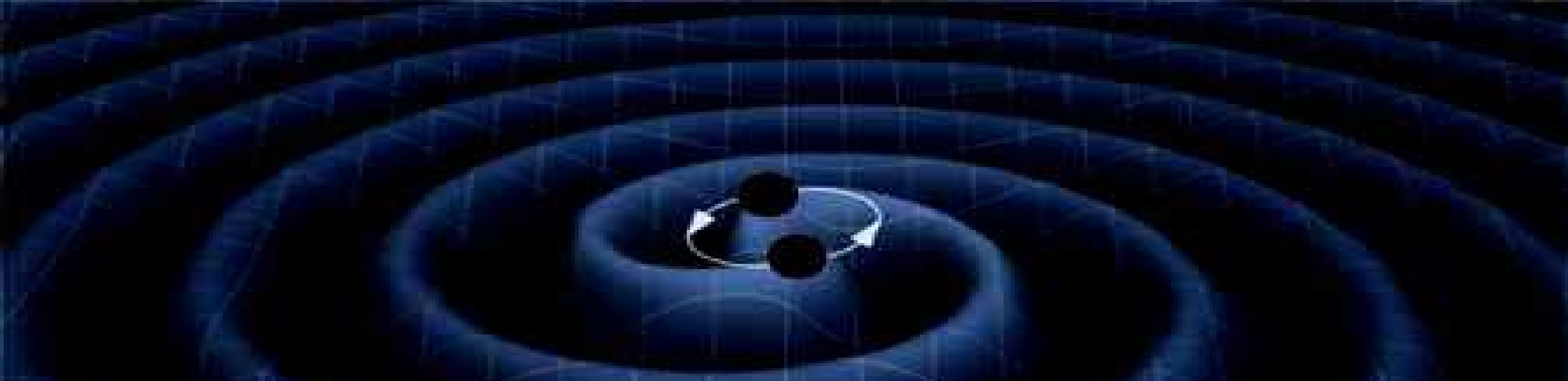


Resolving multiple massive black hole binaries with pulsar timing

Alberto Sesana

Albert Einstein Institute, Golm

(A. Petiteau, S. Babak)

Hannover 6/06/2012

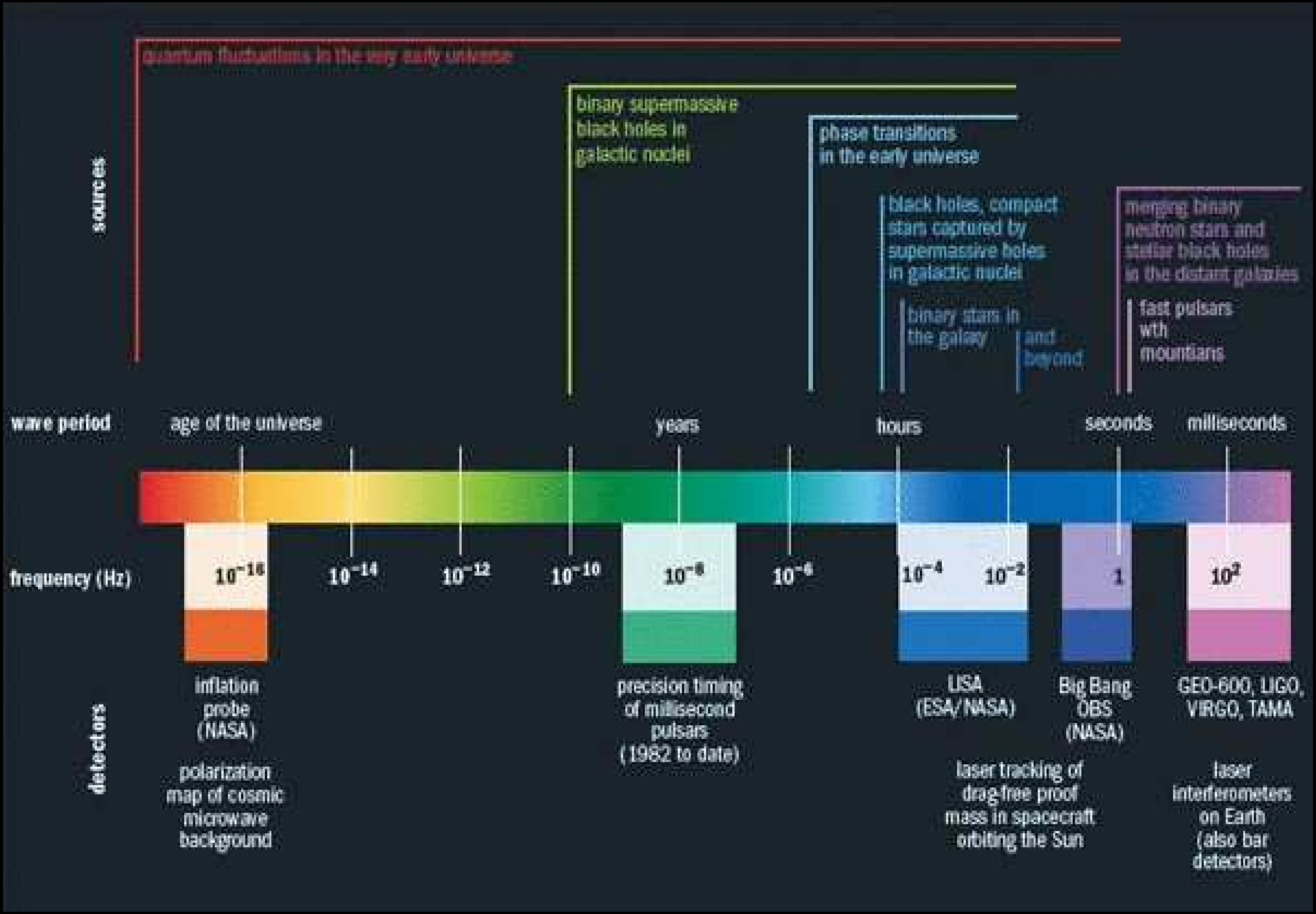
OUTLINE

> Individually resolvable MBHBs

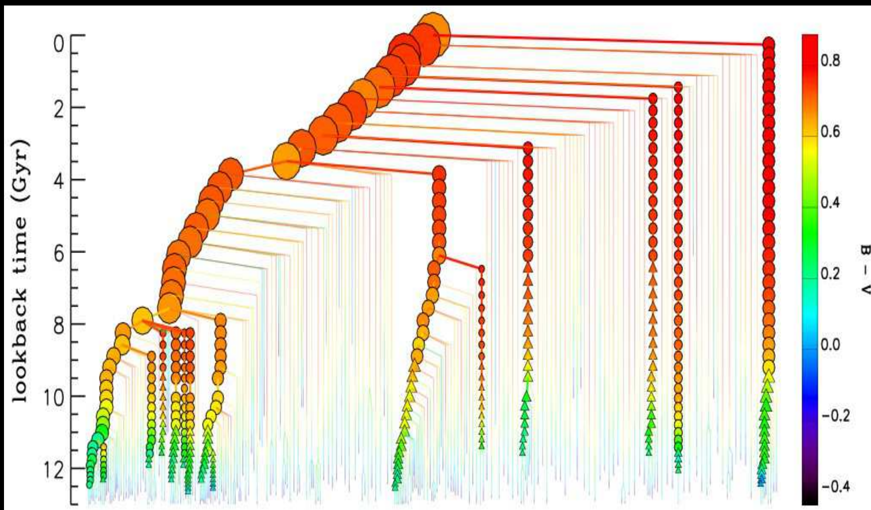
> Detection statistics and search algorithm

> Results and outlook

The gravitational wave spectrum

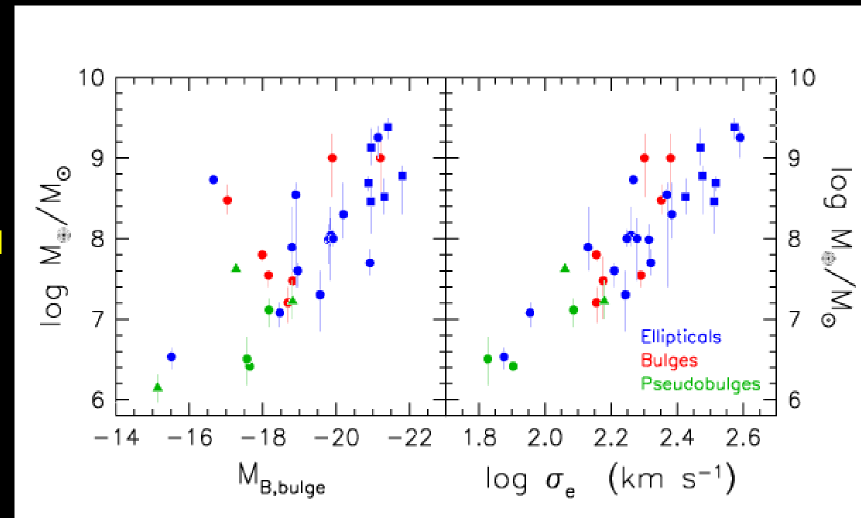


Structure formation in a nutshell



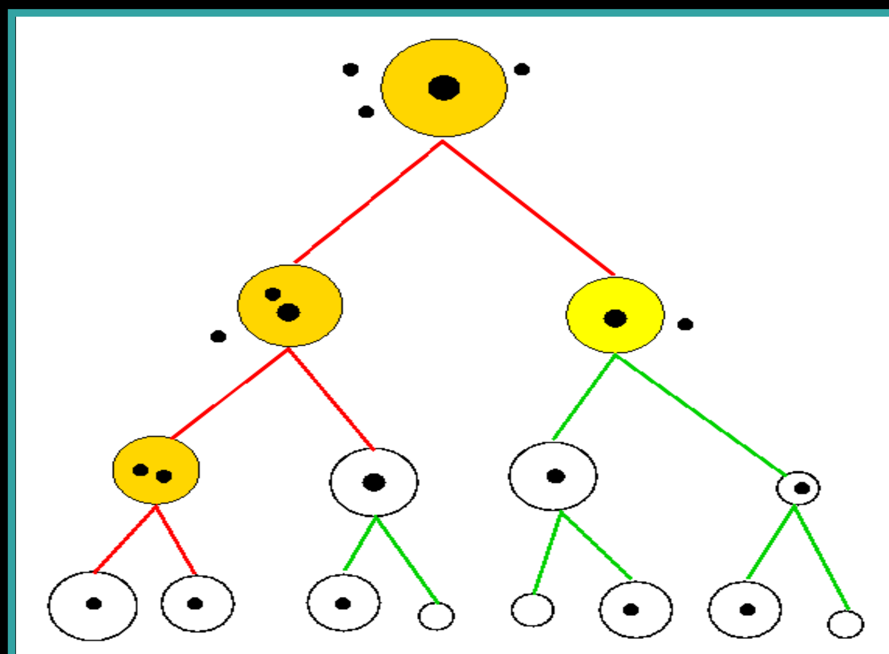
From De Lucia et al 2006

+



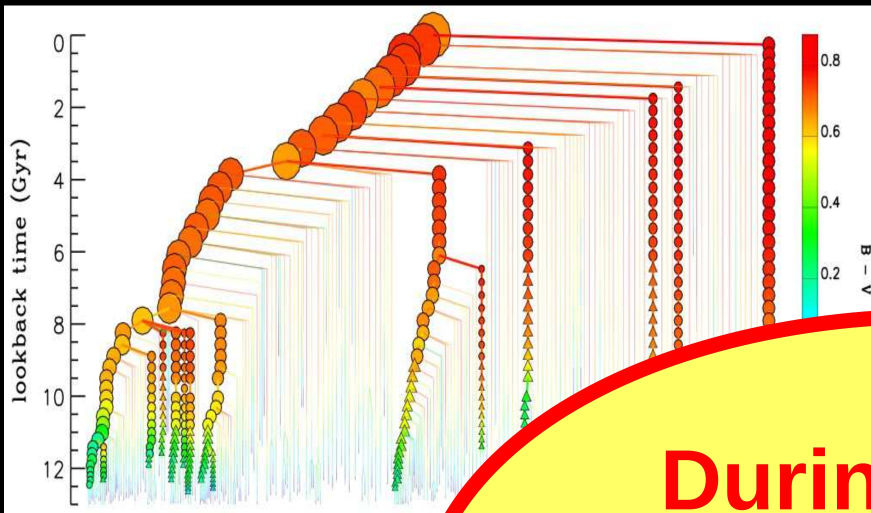
Ferrarese & Merritt 2000, Gebhardt et al. 2000

=

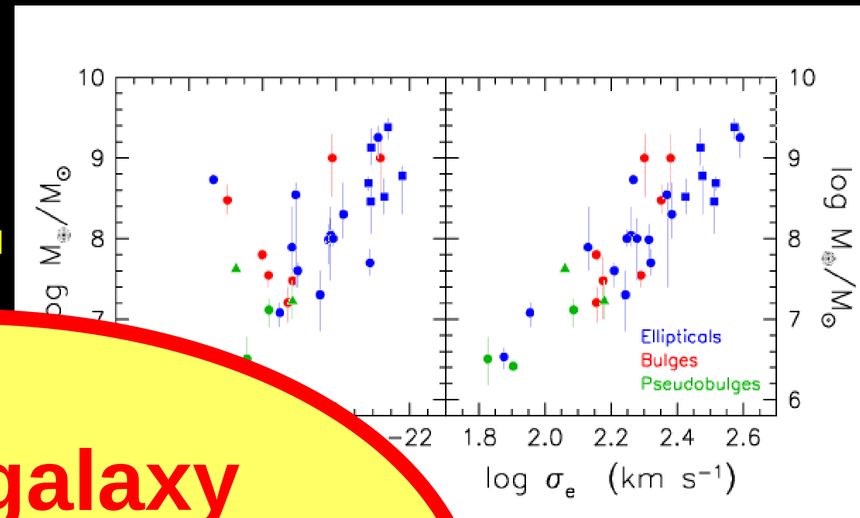


Volonteri Haardt & Madau 2003

Structure formation in a nutshell

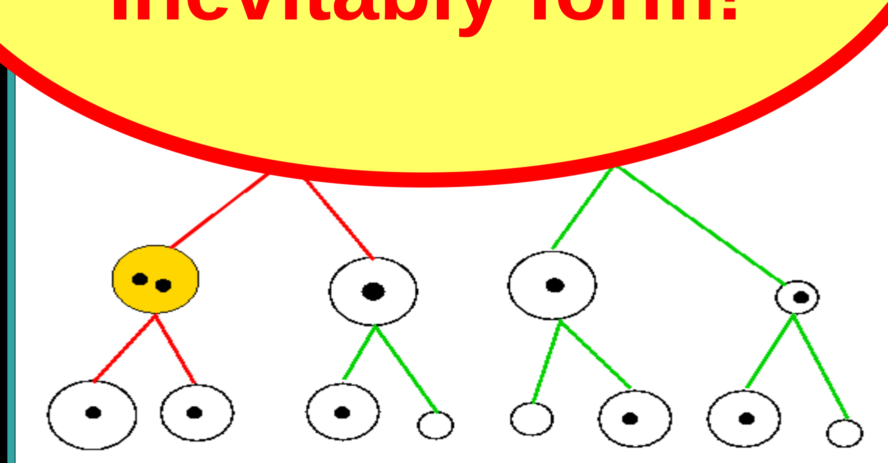


From De Lucia et al 2



Gebhardt et al. 2000

During galaxy mergers, MBHBs will Inevitably form!

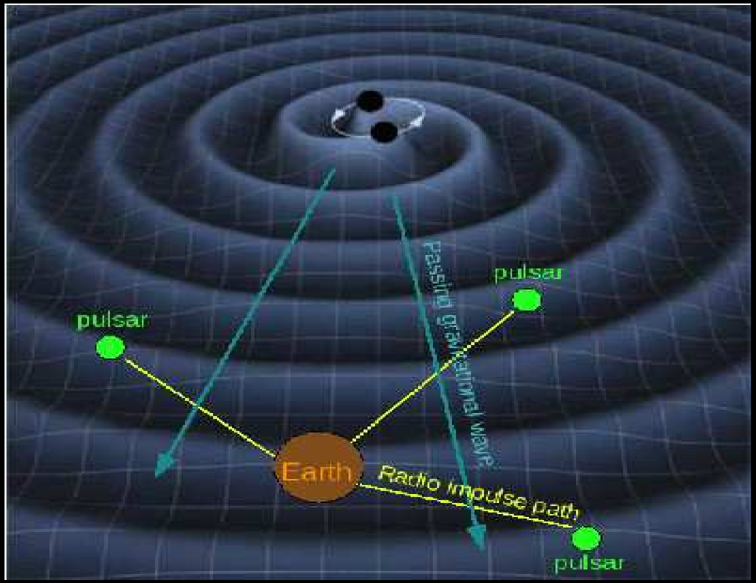


Volonteri Haardt & Madau 2003

The timing residual R

The GW passage cause a modulation of the MSP frequency

$$\frac{\nu(t)-\nu_0}{\nu_0} = \Delta h_{ab}(t) \equiv h_{ab}(t_{\rm p},\hat{\Omega}) - h_{ab}(t_{\rm ssb},\hat{\Omega})$$



The *residual* in the time of arrival of the pulse is the integral of the frequency modulation over time

$$R(t) = \int_0^T \frac{\nu(t)-\nu_0}{\nu_0}\, dt$$

(Sazhin 1979, Helling & Downs 1983, Jenet et al. 2005, Sesana Vecchio & Volonteri 2009)

$$= \frac{\mathcal{M}^{5/3}}{D}\,[\pi f(t)]^{-1/3}$$

$R \sim h/(2\pi f)$

$$\simeq 25.7 \left(\frac{\mathcal{M}}{10^9\, M_\odot}\right)^{5/3} \left(\frac{D}{100\,{\rm Mpc}}\right)^{-1} \times \left(\frac{f}{5\times 10^{-8}\,{\rm Hz}}\right)^{-1/3} {\rm ns}$$

Signal from a MBHB population

Theoretical 'average' spectrum

Contribution of individual sources

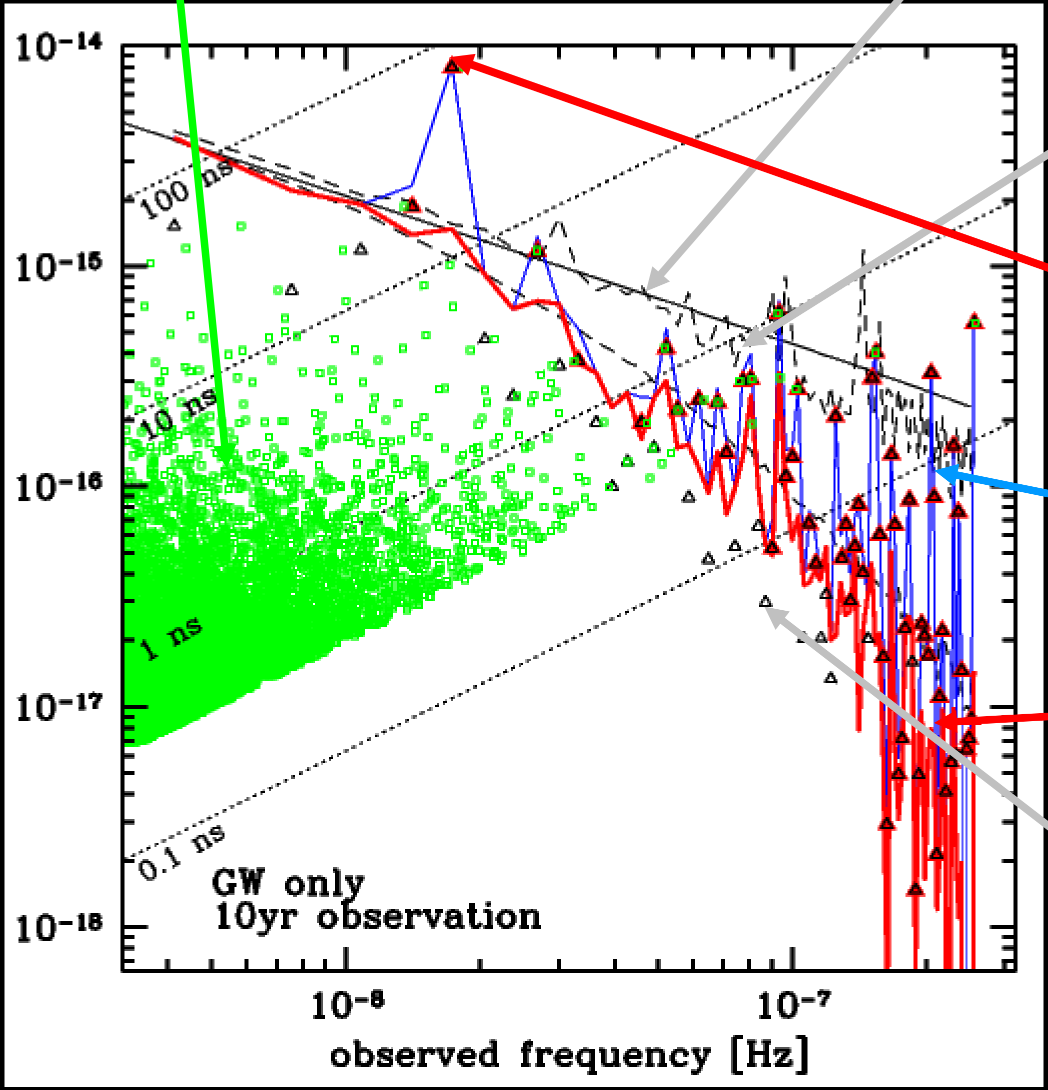
Spectrum averaged over 1000 Monte Carlo realizations

Resolvable systems: **i.e.** ***systems whose signal is larger than the sum of all the other signals falling in their frequency bin***

Total signal

Unresolved background

Brightest sources in each frequency bin



Datasets

We create several datasets

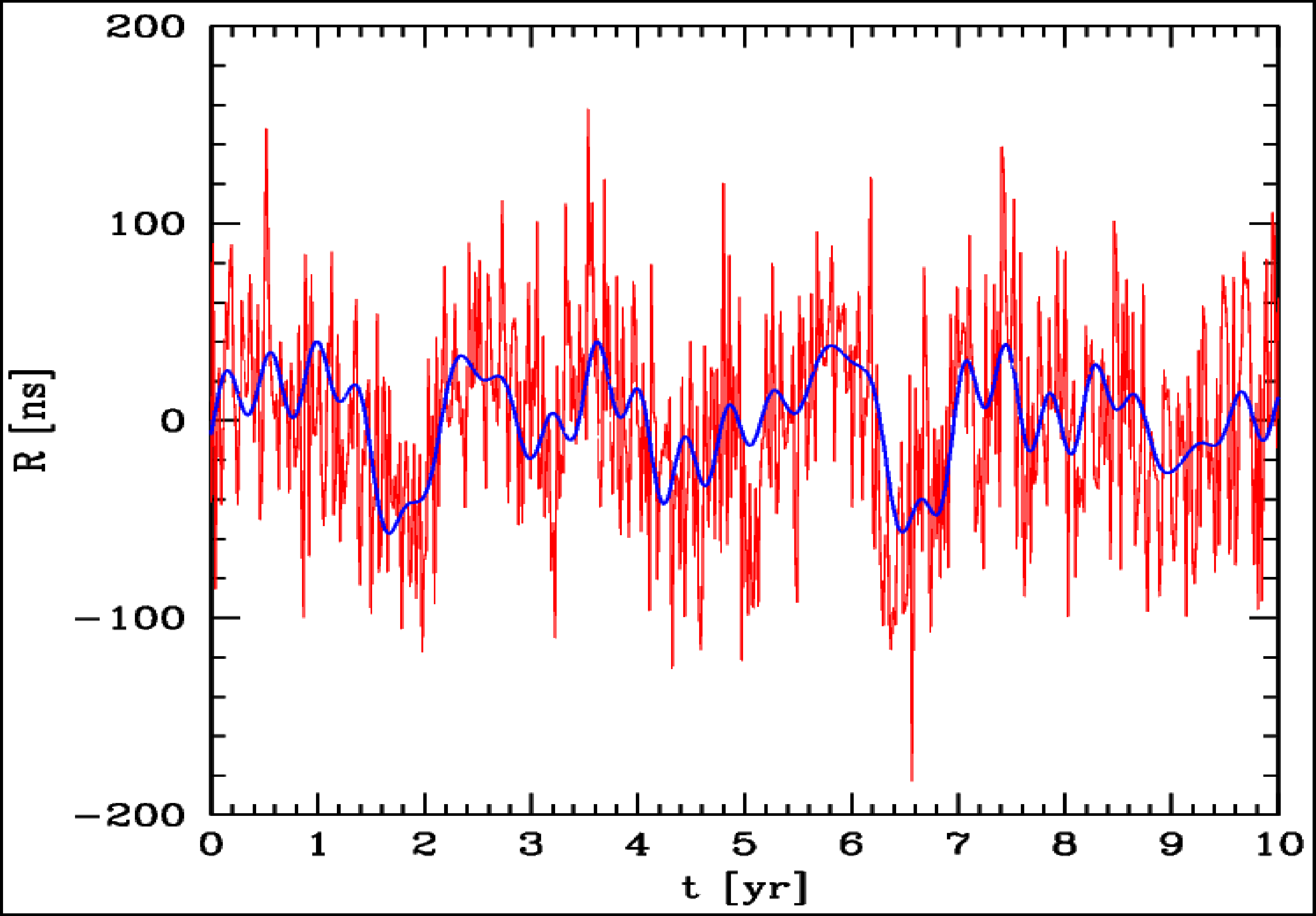
- -Unknown number of sources, random sky location, frequency, etc.
- -Circular non-spinning binaries evolved with PN equations of motion
- -Different number of pulsars in the array (30, 50, 100), random sky location, random distance 1-3 kpc
- -Earth term only and full Earth+pulsar response
- -Different noise levels, equal and unequal noise (white Gaussian)

Test datasets created with increasing level of complexity

- -dataset 1: 5 sources with very high SNR (~50-70)
- -dataset 2: 4 sources, SNR ~10-50, large spread in frequency
- -dataset 3: 8 sources, SNR ~10-30
- -dataset 4: 3 sources, different noise level

Upcoming datasets

- -include full population of the brightest sources coming from state of the art MBH population models
- -SKA-type configuration (100 pulsars, 10ns noise)



200
100
0
−100
−200
R [ns]
0
1
2
3
4
5
6
7
8
9
10
t [yr]

Detection statistics *(Babak & Sesana 2012)*

We model the Earth term only assuming monochromatic non-evolving, non spinning, circular binaries.

The timing residual can be analytically rearranged in the form

$$r_\alpha(t) = \sum_{j=1}^{4} a_{(j)} h_{(j)}^{\alpha}$$

$$a_{(j)} = a_{(j)}(\iota, \psi, \phi_0, \mathcal{A}), \quad h_{(j)}^{\alpha} = h_{(j)}^{\alpha}(t, f, \hat{n}_\alpha, \theta, \phi)$$

We can express the likelihood of a dataset $x_\alpha(t)$ of containing a signal $r_\alpha(t,\lambda)$, where λ is the set of parameters describing the signal, as

$$\log \Lambda_\alpha \sim (x_\alpha || r_\alpha) - \frac{1}{2}(r_\alpha || r_\alpha) \qquad (x||h) \equiv \frac{2}{T_o}\int_0^{T_o} x(t)h(t)dt$$

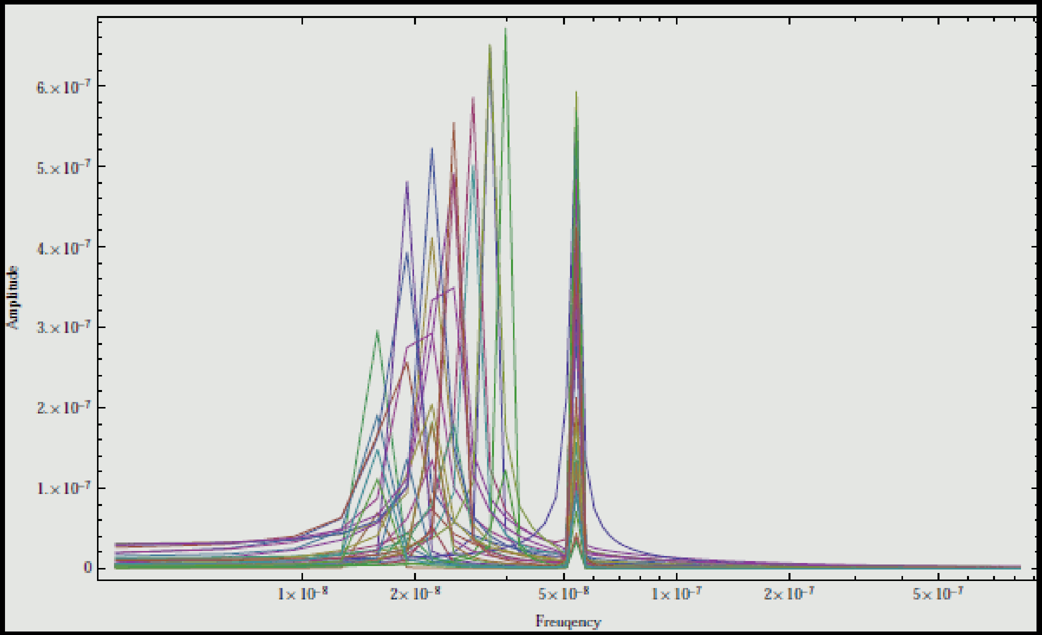
The total likelihood is the sum over all pulsars, which we can write

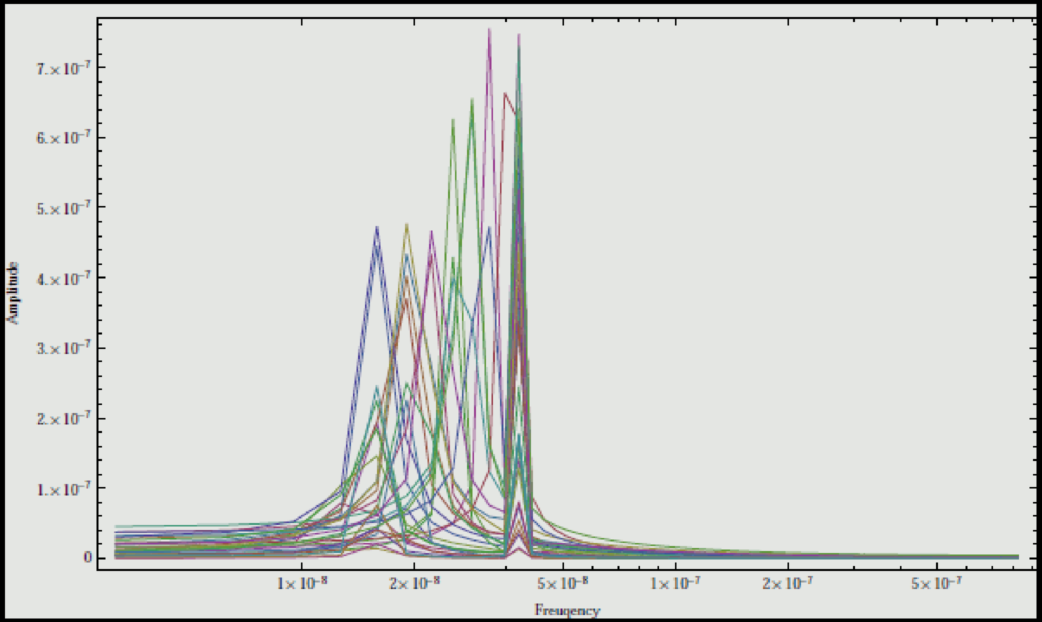
$$\log \Lambda = \sum_{\alpha=1}^{P} \log \Lambda_\alpha \sim \sum_{j=1}^{4} a_{(j)} X_j - \frac{1}{2}\sum_{j=1}^{4}\sum_{k=1}^{4} a_{(j)} a_{(k)} M_{jk}$$

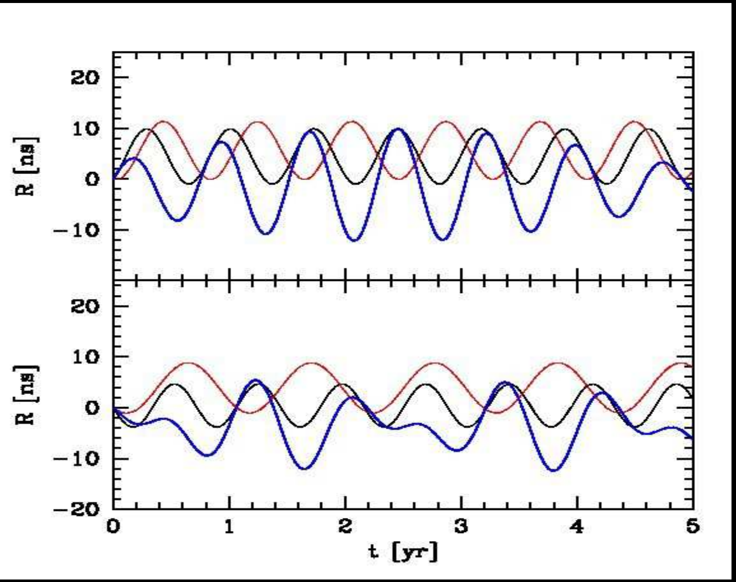
$$X_j \equiv \sum_{\alpha=1}^{P} (x_\alpha || h_{(j)}^{\alpha}), \quad M_{ik} \equiv \sum_{\alpha=1}^{P} (h_{(j)}^{\alpha} || h_{(k)}^{\alpha}).$$

Which allows analytical maximization over a_j. We have 3xN parameters

Why the Earth term only?







> *Earth terms* add up coherently, giving a strong well defined peak

> They are described by 7 parameters only

> *Pulsar terms* fall at different (lower) frequencies and do not add up coherently

> You would need 2x*P* (*P* is the number of pulsars) other parameters to describe all of them

> At this stage we treat them as a source of noise

Search with Genetic Algorithm

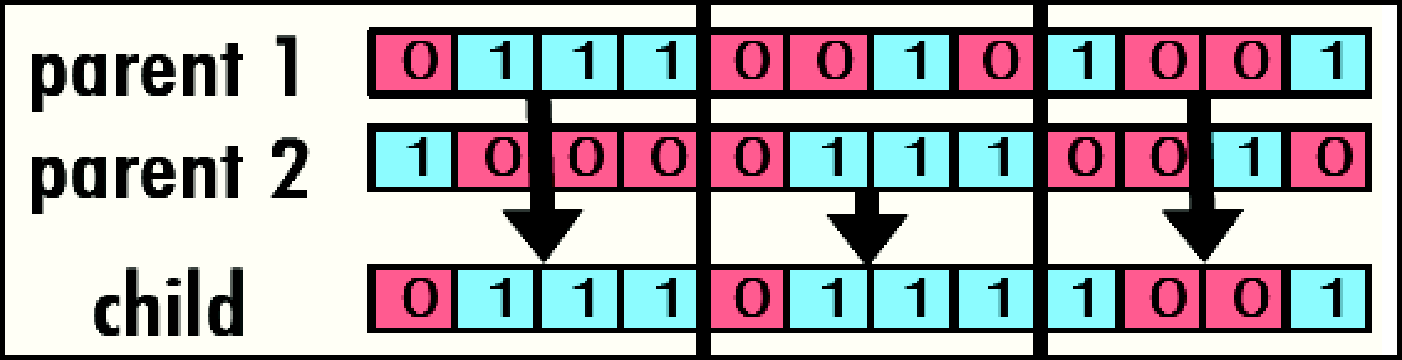
- -Common optimization technique based on natural selection
- -Colony of organisms characterized by “fitness”-likelihood
- -Each organism is described by a set of genes-parameters (N sources, frequencies, sky locations)
- -Strong organisms survive and proliferate

Genetic algorithm		GW search
organism	$\Longleftrightarrow$	template
gene (of an organism)	$\Longleftrightarrow$	parameter (of a template)
allele (of a gene)	$\Longleftrightarrow$	bits (of the value of the parameter)
quality Q	$\Longleftrightarrow$	Maximized Likelihood or A-statistic
colony of organisms	$\Longleftrightarrow$	evolving group of templates
n-th generation	$\Longleftrightarrow$	the state of colony at n-th step of evolution
(selection + breeding) + mutation	$\Longleftrightarrow$	way of exploring the parameter space

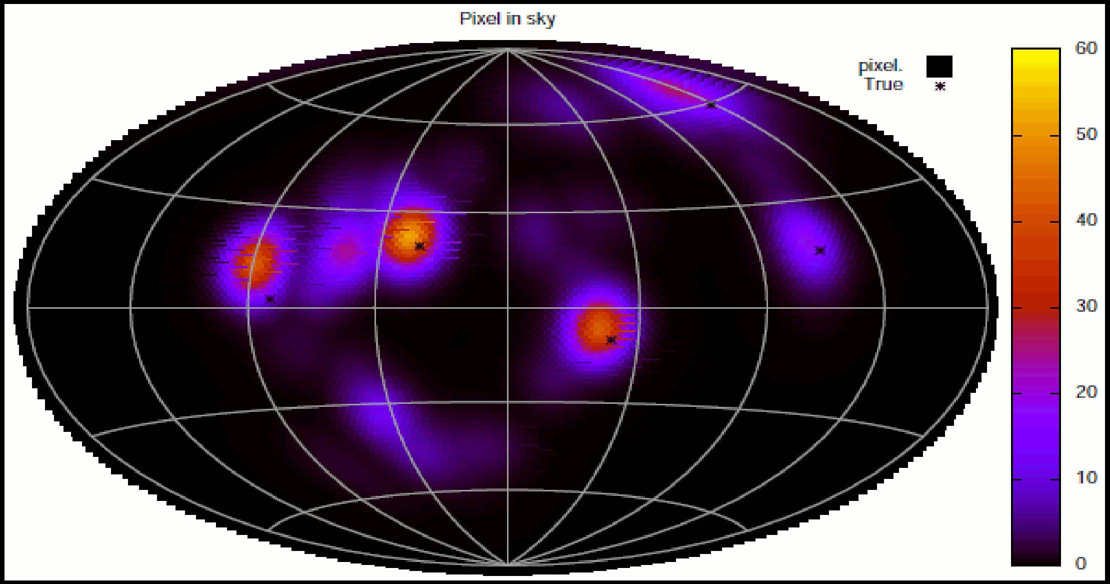
Algorithm scheme

Three basic operations define the evolution of a colony

- ***>Selection***: at any given generation, we select two (or three) parents for breeding. Based on the quality of each organism: the higher the likelihood of an organism, the higher is the probability to be chosen
- ***>Breeding***: the genes of the two parents are combined to produce a child (many possible ways to do that)
- ***>Mutation***: some of the genes are mutated randomly according to some probability

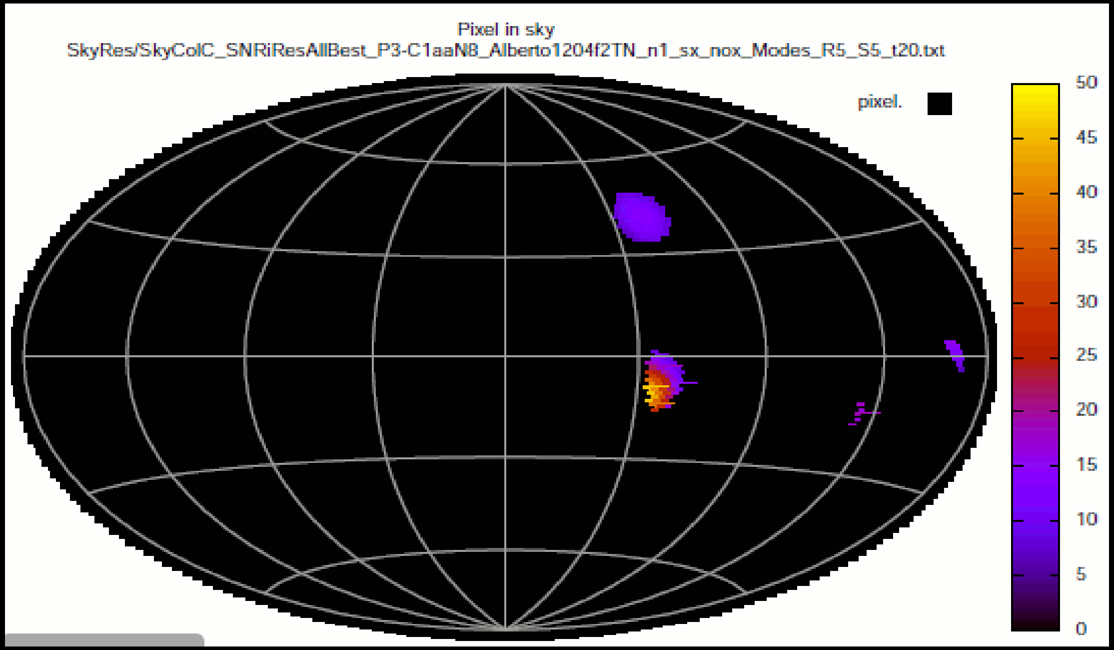


Preliminarily Results



Dataset 1

We find all sources but:
-secondary maxima
-bias in sky location

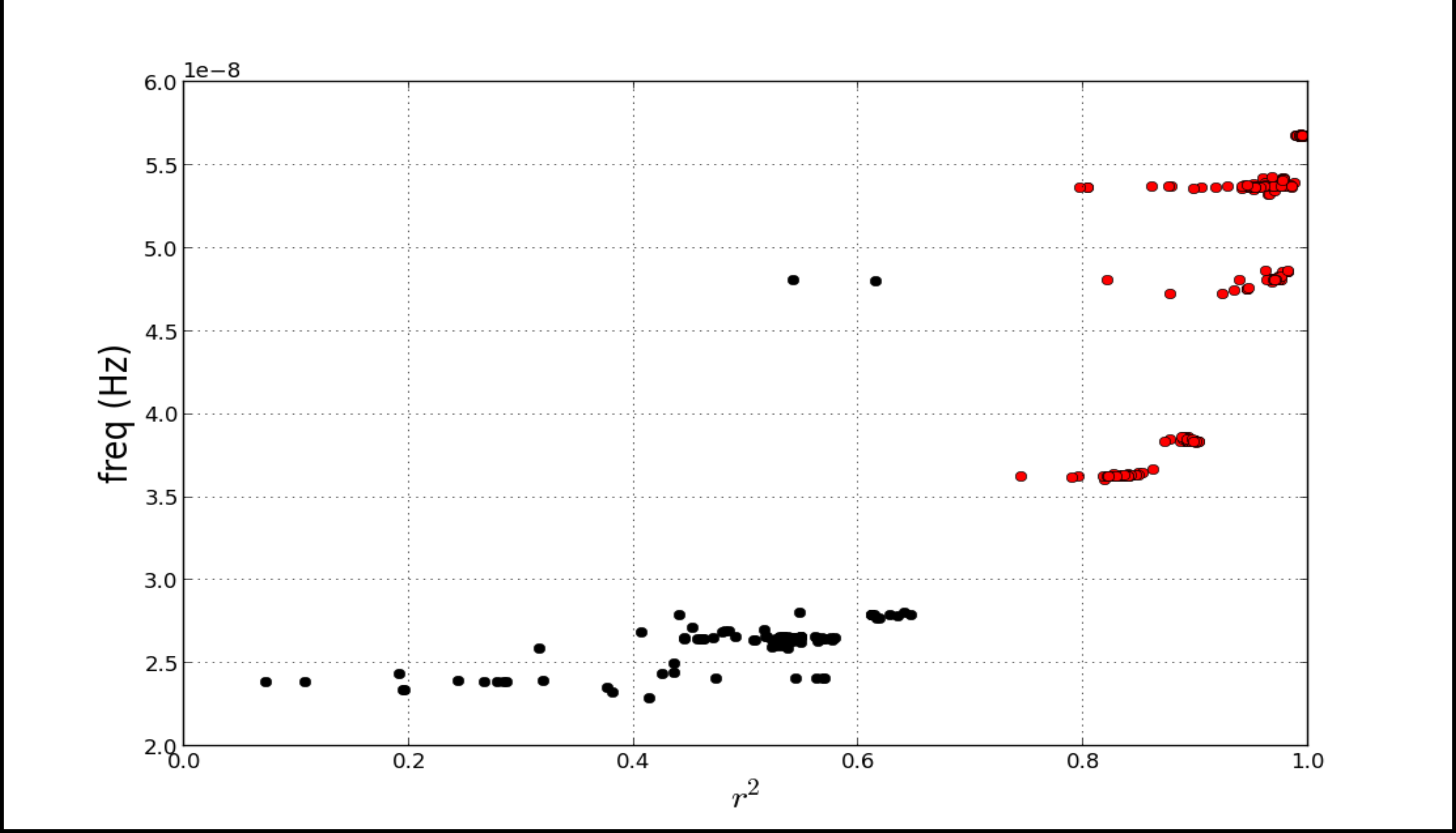


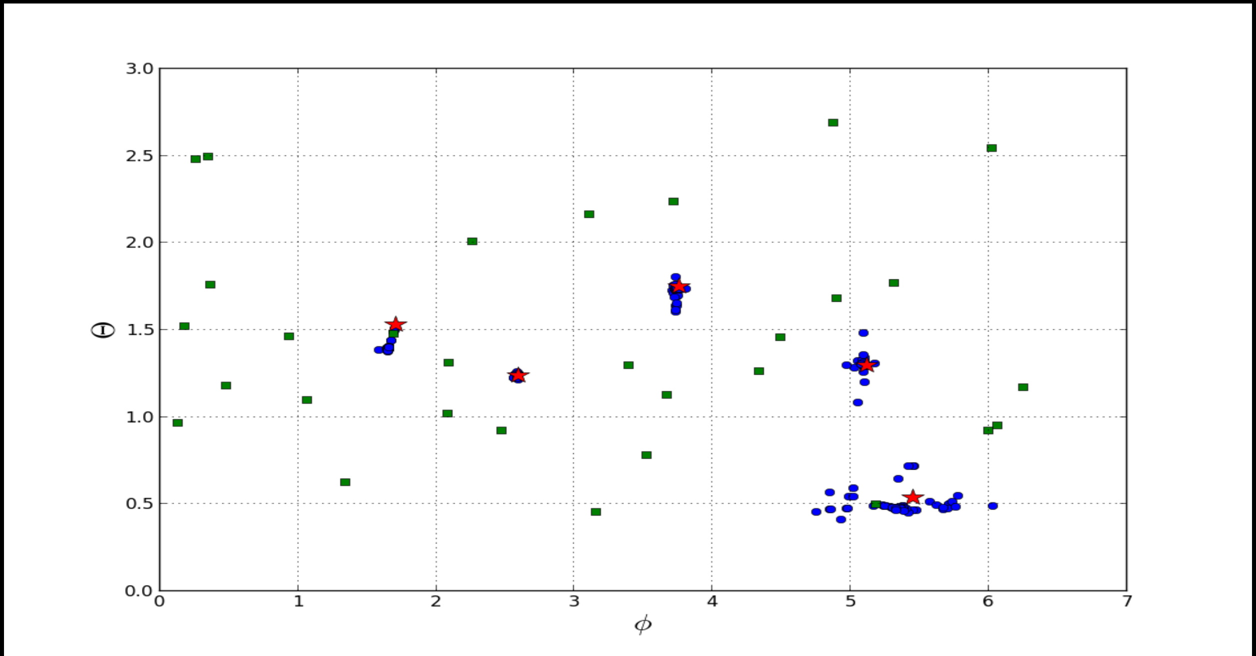
Dataset 2

We find only two sources:
-secondary maxima
-bias in sky location
-high frequency sources missing

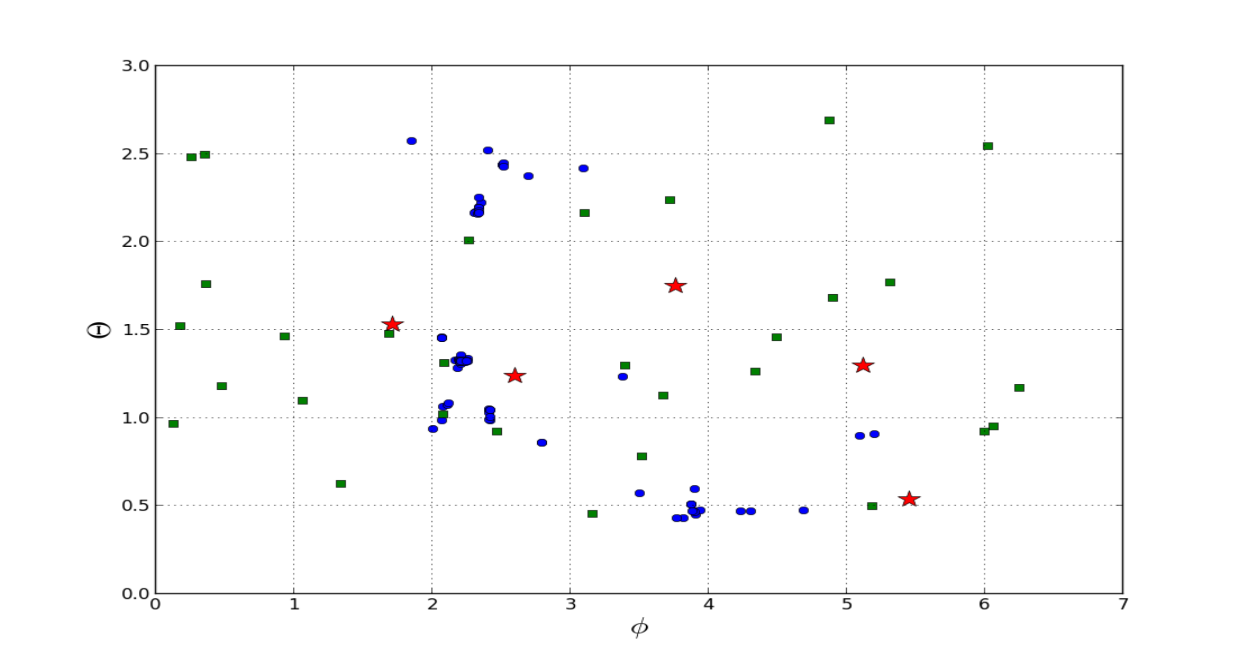
1-Curing false positives: correlation analysis

- -When we know the sky location of a candidate source, we now exactly theoretically how each pulsar should contribute to the SNR of the source 'if it were an Earth term'
- -We correlate theoretical expectations to what actually happens in the data



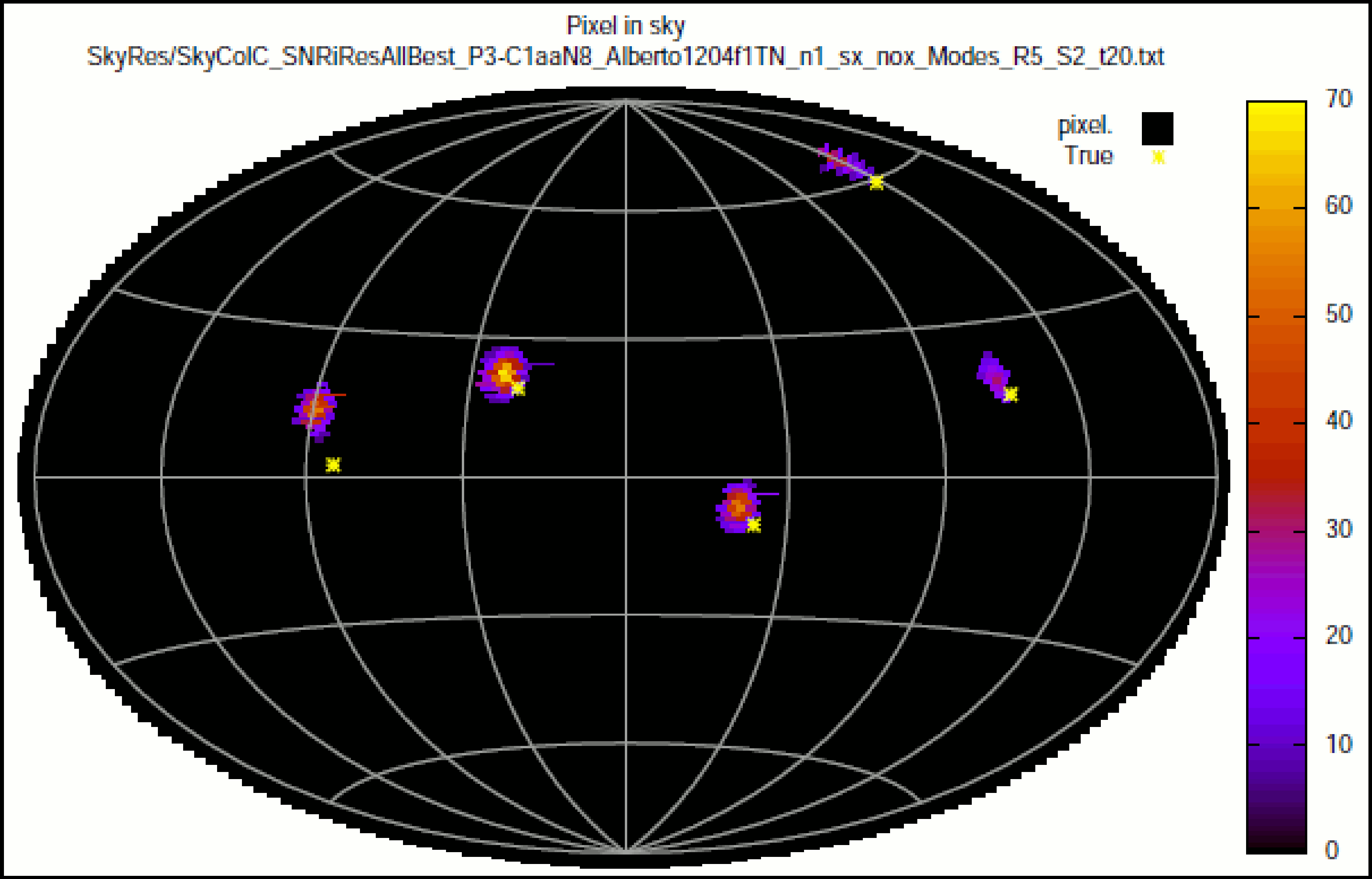


The modes showing high correlation are those with higher SNR and fall around the true source locations.



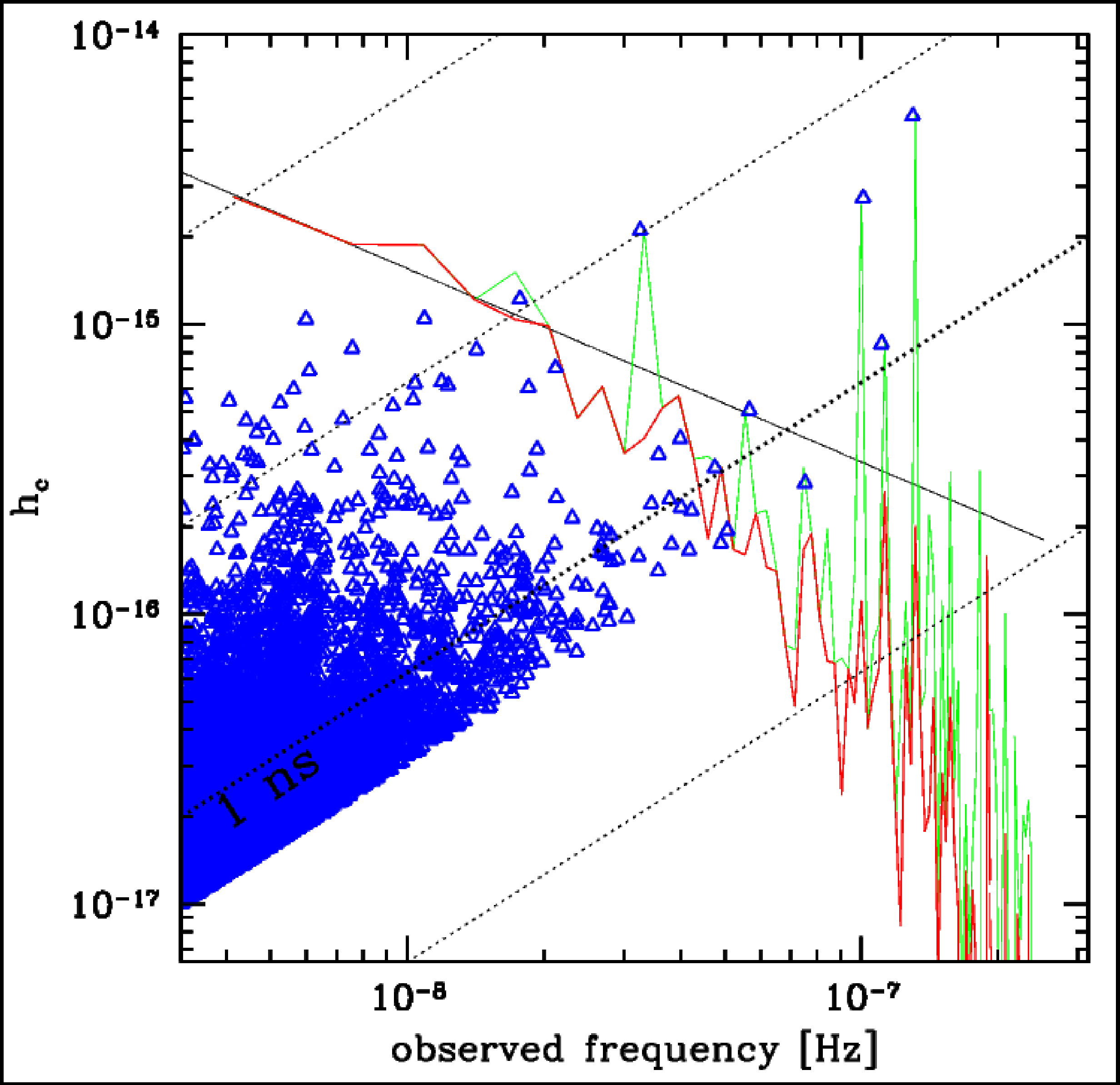
The modes showing low correlation are those with lower SNR and are scattered around

Dataset 1 “preliminarily final” results



We get rid of all the 'false' positives
Some bias is still present in the final answer

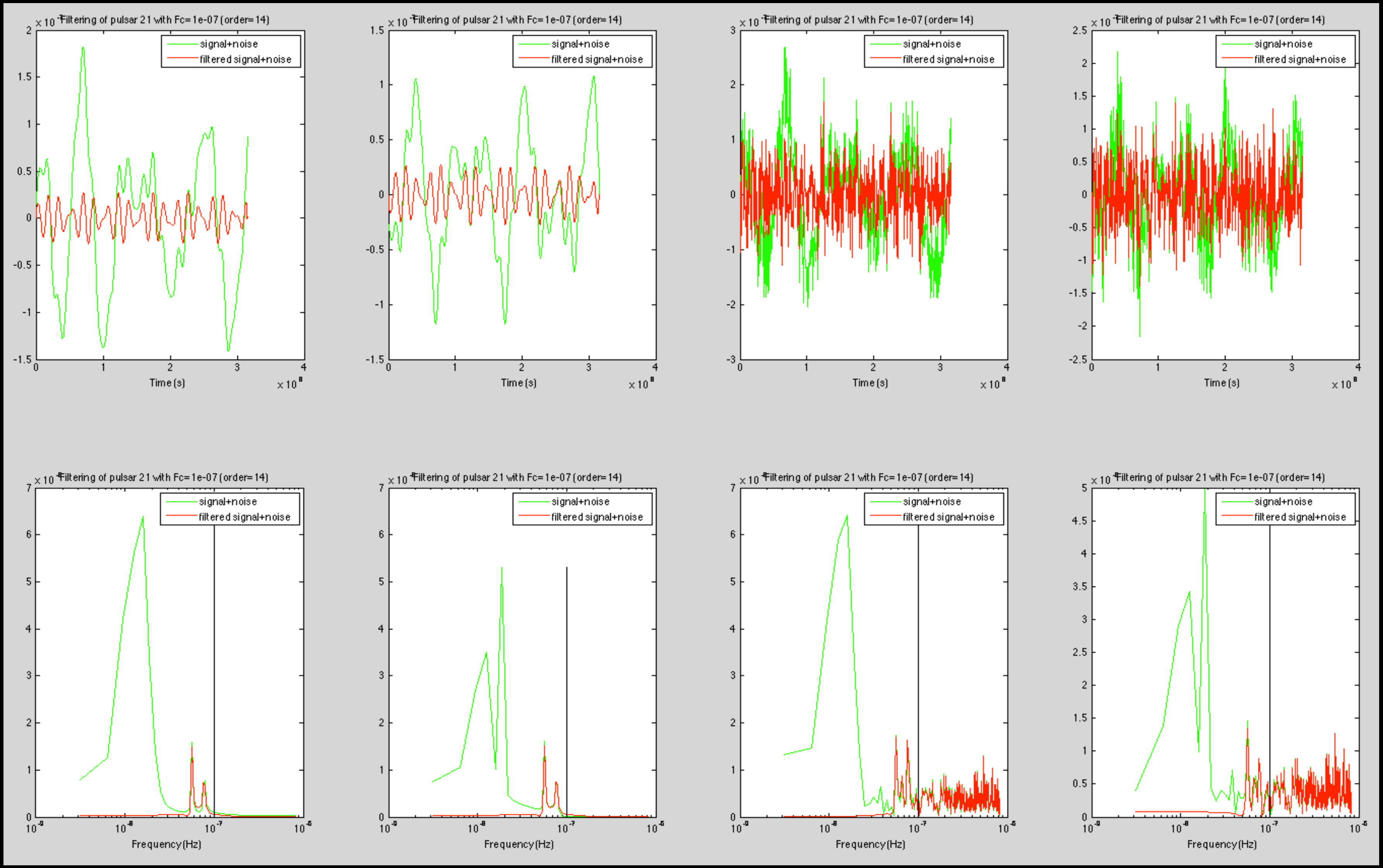
2-Finding high frequency sources: filtering



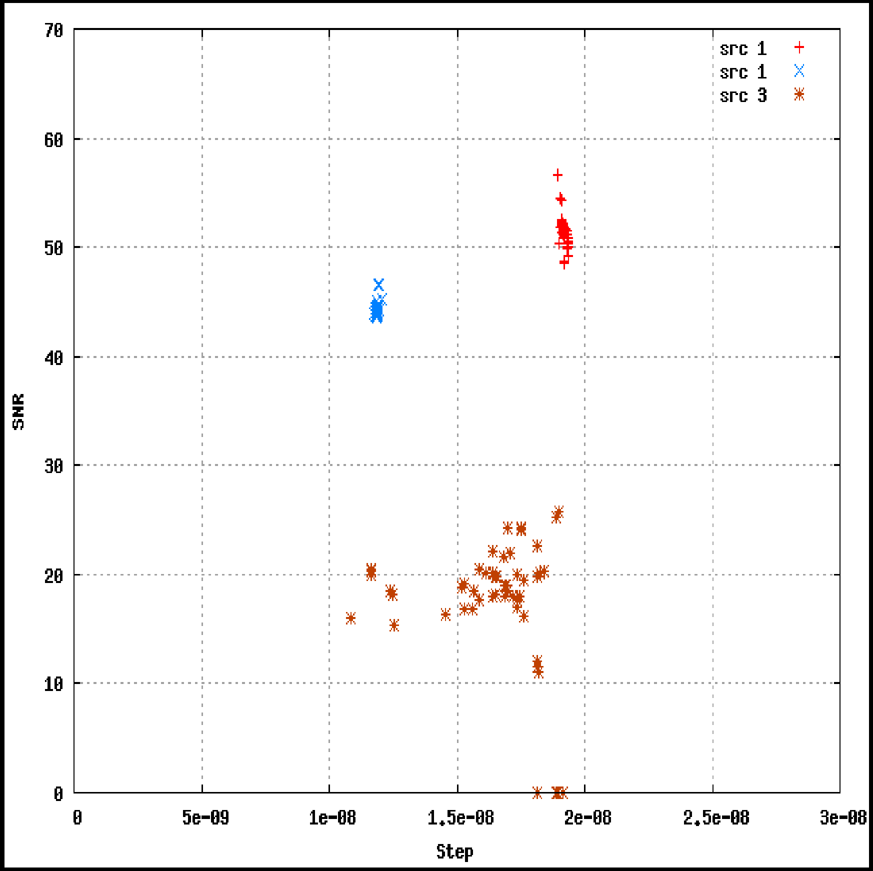
Low frequency sources are much brighter than high frequency ones.

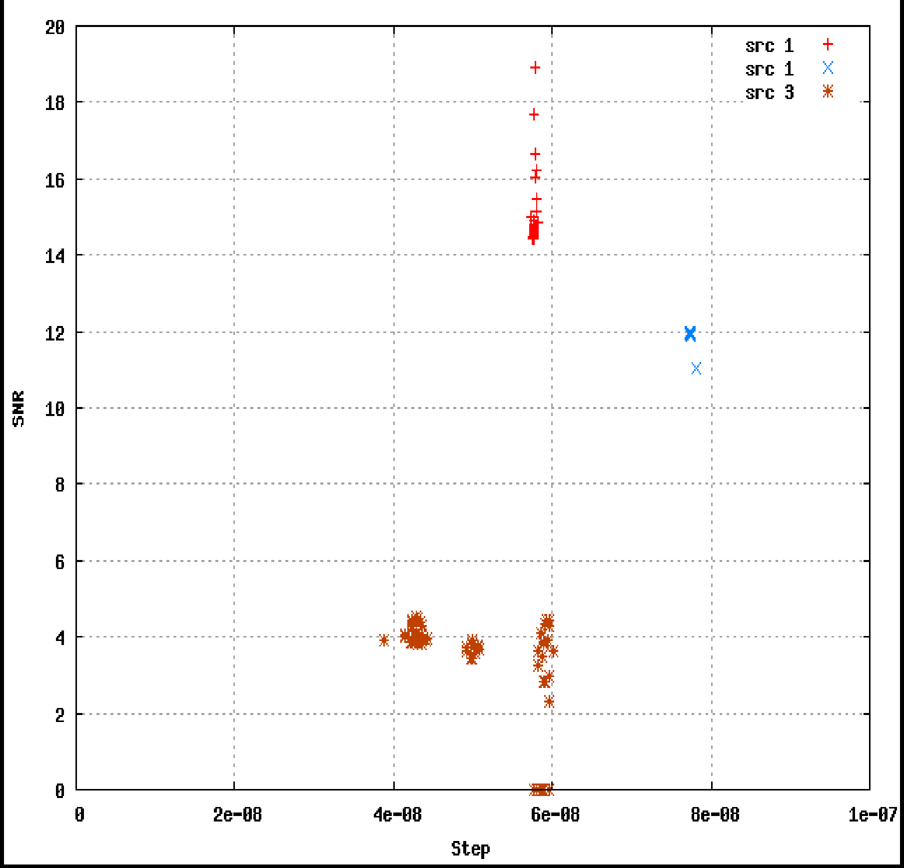
The detection algorithm prefers low frequency signals and gets stuck on high SNR combinations of pulsar terms

We apply a *high pass filter* to the data, and we progressively include lower and lower frequencies

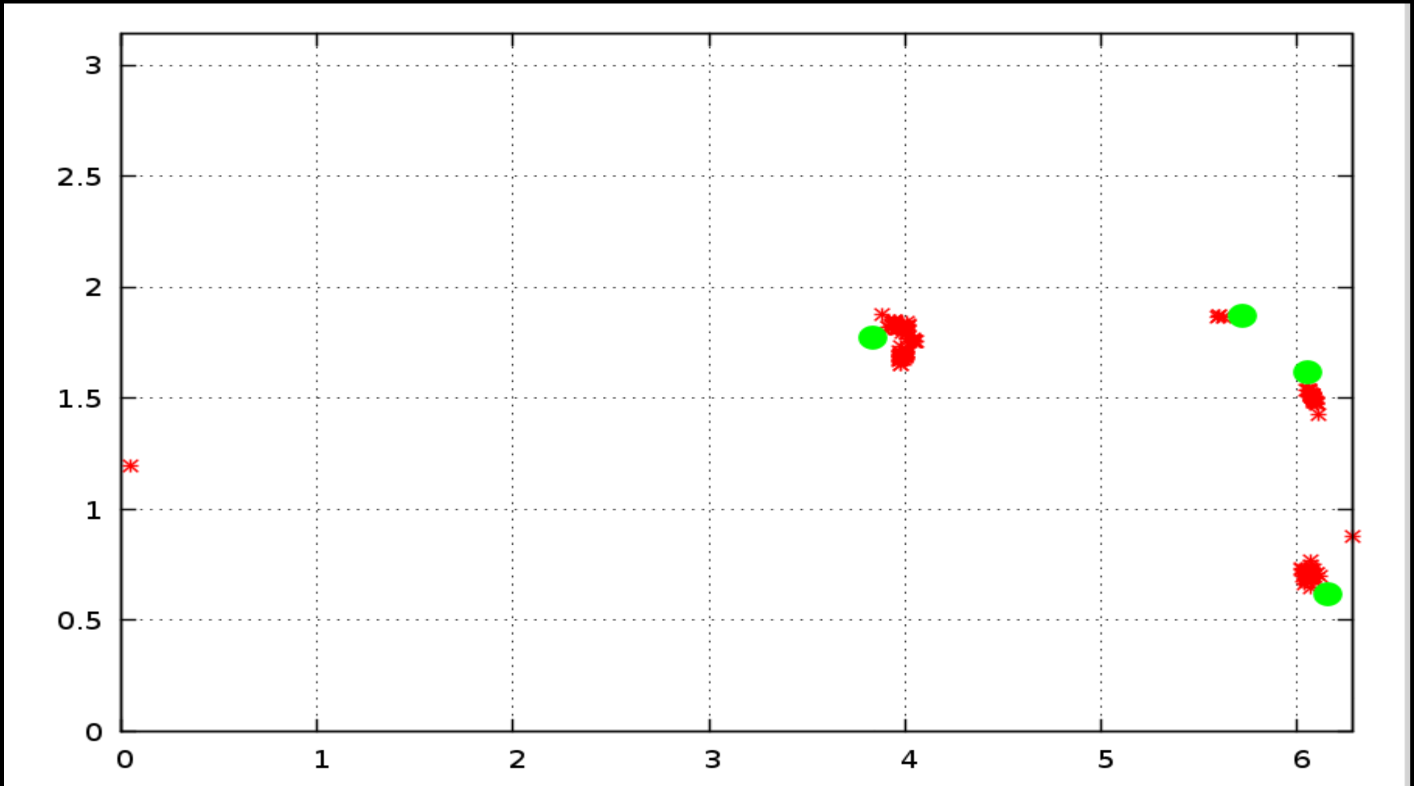


Dataset 2 “preliminarily final” results





The filtering procedure allows us to *clearly identify two high frequency sources* whose SNR is actually lower than the spurious combinations of low frequencies pulsar terms



After the filtering procedure *we correctly identify all the four injected sources at the approximately correct sky location*.

-sky location correct within few degrees (some bias still present)
-frequency correct at the Fourier frequency bin level
-Approximately correct SNR

Outlook

-Further testing on progressively more realistic datasets (drawn from realistic population models, including more sources, eccentricity)

-Extension to unevenly sampled data

-Inclusion of more complex forms of noise

-This is just a search algorithm! Coherent implementation of a full pipeline (MCMC to explore the maxima, give errors on parameters, etc.)

-Integration of the algorithm in the EPTA data analysis pipeline

-Open questions: bias? Pulsar term inclusion?