$\mathcal{F}$-Statistic Search on the Second Mock LISA Data Challenge

Reinhard Prix and John T. Whelan

Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Potsdam, Germany

reinhard.prix@aei.mpg.de, john.whelan@aei.mpg.de

Abstract

*The $\mathcal{F}$-statistic is an optimal detection statistic for continuous gravitational waves, i.e., long-duration (quasi-) monochromatic signals with slowly-varying intrinsic frequency. This method was originally developed in the context of ground-based detectors, but it is equally applicable to LISA where many signals fall into this class of signals. We report on the application of a LIGO/GEO $\mathcal{F}$-statistic code to LISA using the **long-wavelength limit** (LWL) approximation, and we present results of our search for white-dwarf binary signals on the second Mock LISA Data Challenge (MLDC2).*

$\mathcal{F}$-Statistic Method

A white-dwarf binary GW signal $s(t)$ with frequency f, located at ecliptic latitude β and longitude λ is characterized by its *Doppler parameters* $\theta = \{f, \beta, \lambda\}$ and *amplitude parameters* $\mathcal{A} = \{h_0, \cos\iota, \psi, \phi_0\}$. A suitable reparametrization $\{\mathcal{A}^\mu\}_{\mu=1}^4$ of the amplitude parameters allows the factorization

$$s(t; \mathcal{A}, \theta) = \mathcal{A}^\mu \, h_\mu(t; \theta) \,, \qquad (1)$$

with automatic summation over $\mu = 1, \ldots 4$. The optimal detection statistic is the *likelihood ratio* Λ:

$$\ln \Lambda(x; \mathcal{A}, \theta) \equiv (x \| s) - \frac{1}{2}(s \| s) \,, \qquad (2)$$

where $x(t)$ is the measured strain data from the detector, and $(.\|.)$ is the (Wiener) matched-filtering scalar product. By analytically maximizing $\ln\Lambda$ over the unknown $\mathcal{A}^\mu$, we can find their maximum-likelihood estimators (MLE) as

$$\mathcal{A}^\mu_{\rm cand}(x; \theta) = \mathcal{M}^{\mu\nu} (x \| h_\nu) \,, \qquad (3)$$

where $\mathcal{M}^{\mu\nu}$ is the matrix inverse of $\mathcal{M}_{\mu\nu} \equiv (h_\mu \| h_\nu)$. Substituting the MLEs $\mathcal{A}^\mu_{\rm cand}$ into the log-likelihood (2), we obtain the $\mathcal{F}$-statistic, which only depends on the unknown Doppler parameters θ:

$$2\mathcal{F}(x; \theta) \equiv |\mathcal{A}_{\rm cand}|^2 \,, \qquad (4)$$

where the norm of amplitude 4-vectors $\mathcal{A}$ is defined as

$$|\mathcal{A}|^2 \equiv \mathcal{A}^\mu \, \mathcal{M}_{\mu\nu} \, \mathcal{A}^\nu \,. \qquad (5)$$

☞ Using the $\mathcal{F}$-statistic, we only need to search over the unknown Doppler-parameters $\theta = \{f, \beta, \lambda\}$. The $\mathcal{F}$-statistic follows a (non-central) χ^2-distribution with four degrees of freedom. When targeting an exact signal ("key") at position $\theta = \theta_{\rm key}$ and amplitudes $\mathcal{A}_{\rm key}$, the χ^2 non-centrality is $|\mathcal{A}_{\rm key}|^2$, and the expectation-value $E[2\mathcal{F}]$ is

$$E\left[2\mathcal{F}(x; \theta_{\rm key})\right] = 4 + |\mathcal{A}_{\rm key}|^2 \,, \qquad (6)$$

where $|\mathcal{A}_{\rm key}| = \sqrt{(s\|s)}$ is the "signal-to-noise ratio" (SNR).

MLDC2 Pipeline

Challenge 2.1 of the MLDC2 included a population of 26 million galactic WD binaries. Of those, 59401 were designated as "bright" sources, the norms of whose amplitude parameter vectors are shown in Fig. 1.

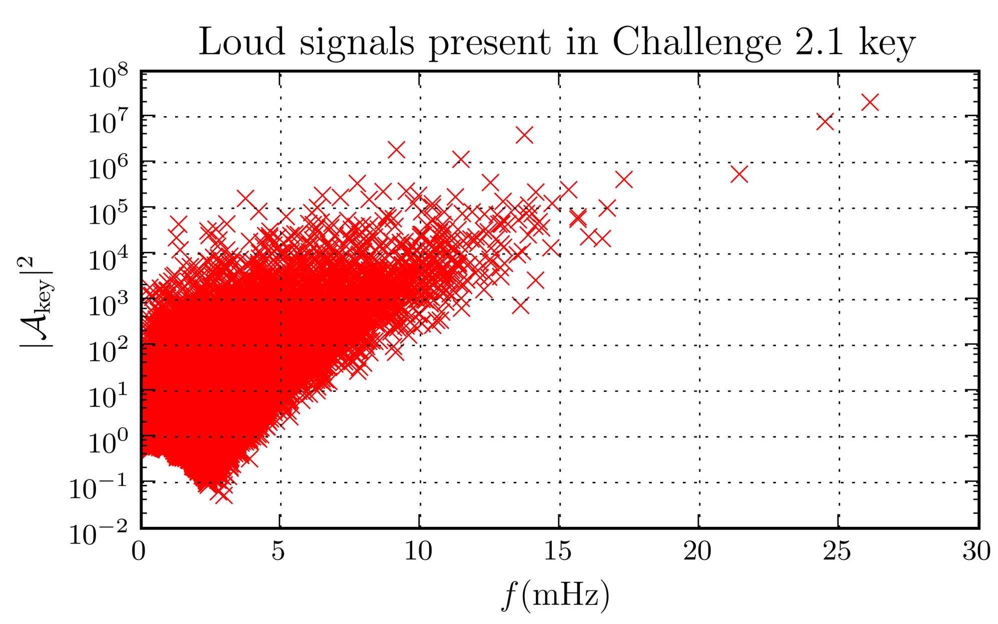

Figure 1: *Sources in Challenge 2.1 "loudest" key file: norms $|\mathcal{A}_{\rm key}|^2$ (corresponding to SNR^2) of injected amplitude-vectors $\mathcal{A}_{\rm key}$ as function of frequency f.*

The large number of detectable sources makes it difficult to distinguish the actual ("primary") from "secondary" maxima of the detection statistic $2\mathcal{F}(x;\theta)$ in Doppler parameter space. Our pipeline is based on the empirical observation that primary maxima show better coïncidence between different TDI variables X, Y, Z than secondary maxima. The coïncidence criterion is based on the *metric* g_{ij} in Doppler parameter space, namely

$$m = g_{ij} \, d\theta^i \, d\theta^j + \mathcal{O}(d\theta^3) \,, \qquad (7)$$

which attributes a "distance" m to the Doppler offsets $d\theta = \{df, d\beta, d\lambda\}$.

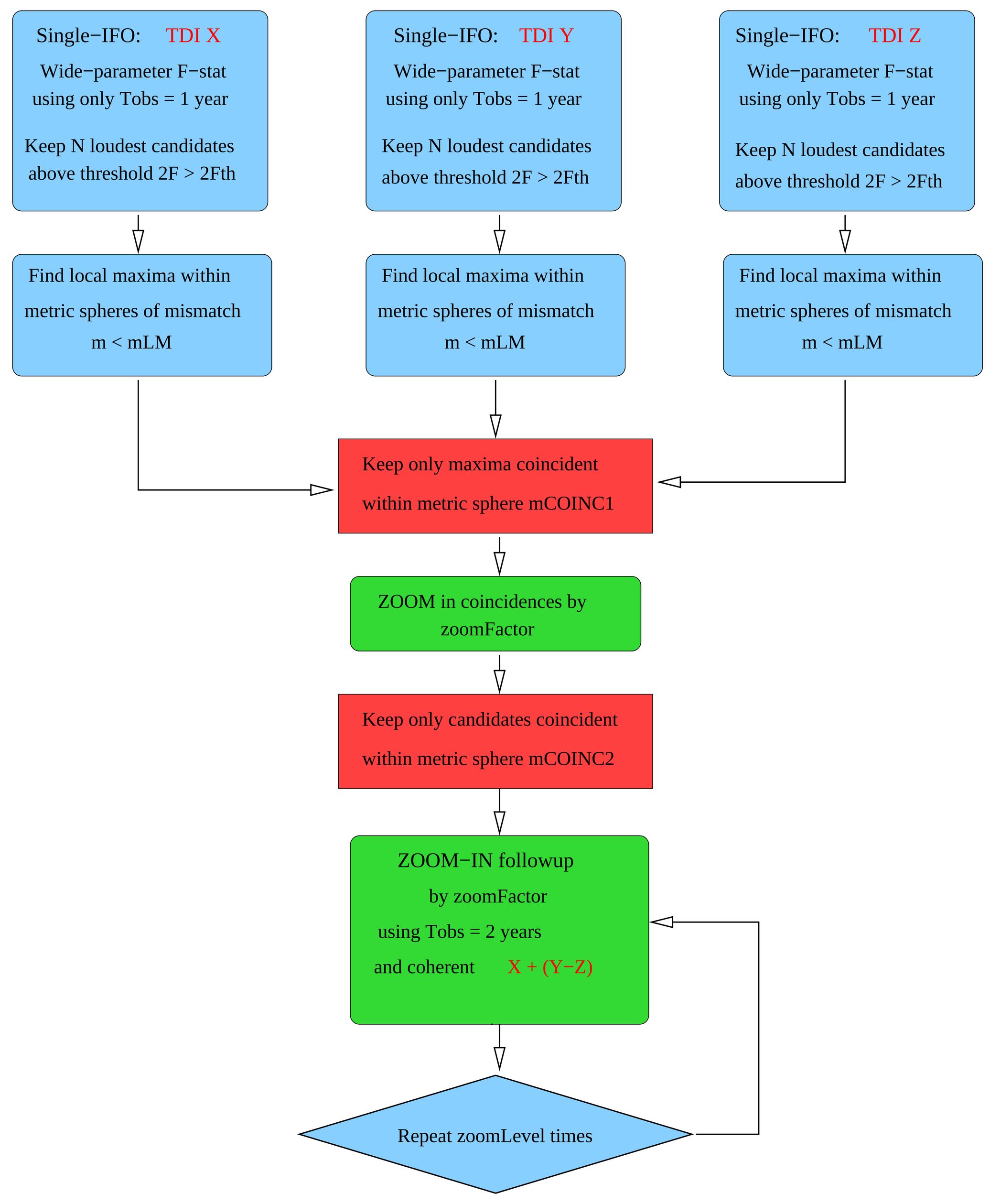

Figure 2: *Pipeline used in the MLDC2 search. The coïncidence criterion used to distinguish primary from secondary maxima was $m_{\rm COINC2} \le 0.2$.*

Results

Our pipeline identified 1777 candidate signals, with $2\mathcal{F}$ values ranging from 44.8 to 1.02×10^6. Of the 59401 "bright" sources in the key, 6872 had $|\mathcal{A}_{\rm key}|^2 > 40$. To evaluate our results, we identified each candidate with the loudest "bright" source within a Doppler mismatch of $m \le 1$. If there was no "bright" source within that Doppler window, we considered the candidate to be a false alarm. If there were multiple bright sources in the Doppler window, the ones with lower $|\mathcal{A}_{\rm key}|$ values were effectively "hidden" from association with a candidate. The results of this identification are illustrated in Fig. 3 and summarized in Table 1.

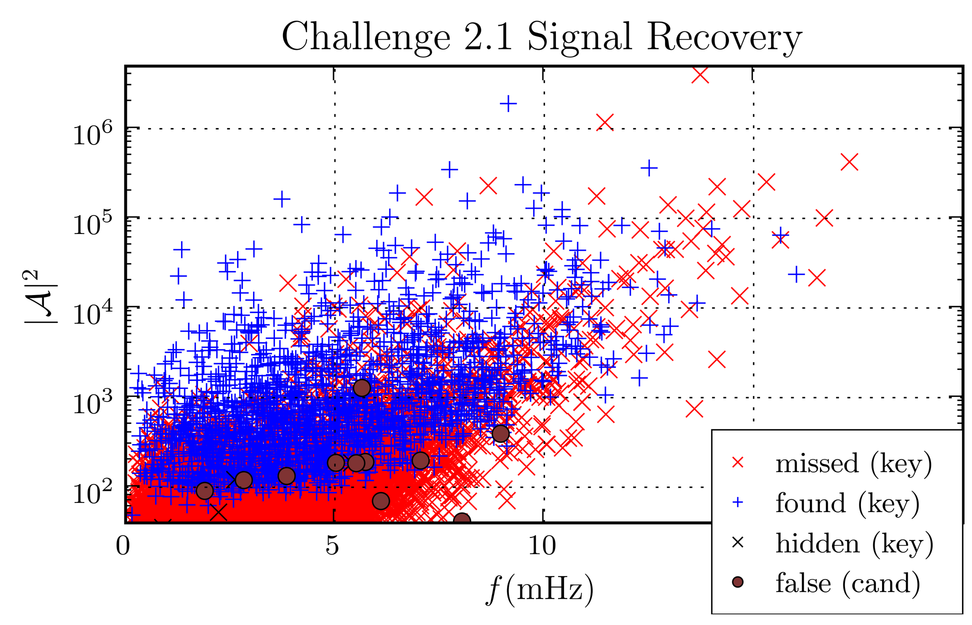

Figure 3: *Overview of signal recovery in Challenge 2.1.*

Freqs	Found	Missed	Hidden	False
0–5 mHz	1012	3642	2	3
5–10 mHz	679	1363	1	8
10–15 mHz	73	90	0	0
15–20 mHz	2	5	0	0
20–27 mHz	0	3	0	0

Table 1: *Signals found in Challenge 2.1, along with missed sources, sources which were "hidden" by coïncidence with louder ones, and false alarms. Only sources with $|\mathcal{A}_{\rm key}|^2 > 40$ are included in the "missed" and "hidden" categories.*

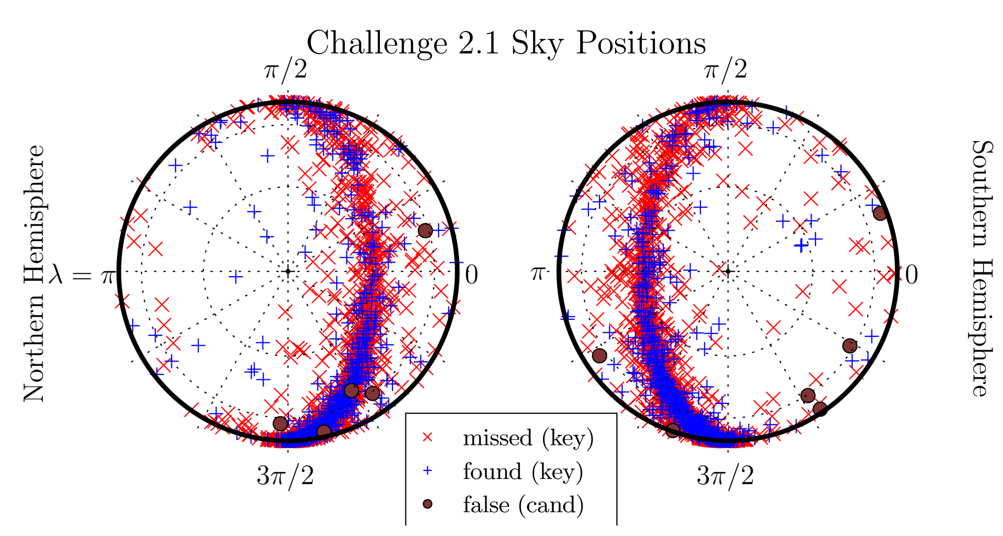

Figure 4: *Sky positions of recovered and missed signals. The galactic plane can be clearly seen as a great circle containing most of the signals and sources.*

As already seen in MLDC1 [1], the use of the LWL affects the estimation of the amplitude parameters (Fig. 6), but does not strongly affect the Doppler localization (Fig. 5).

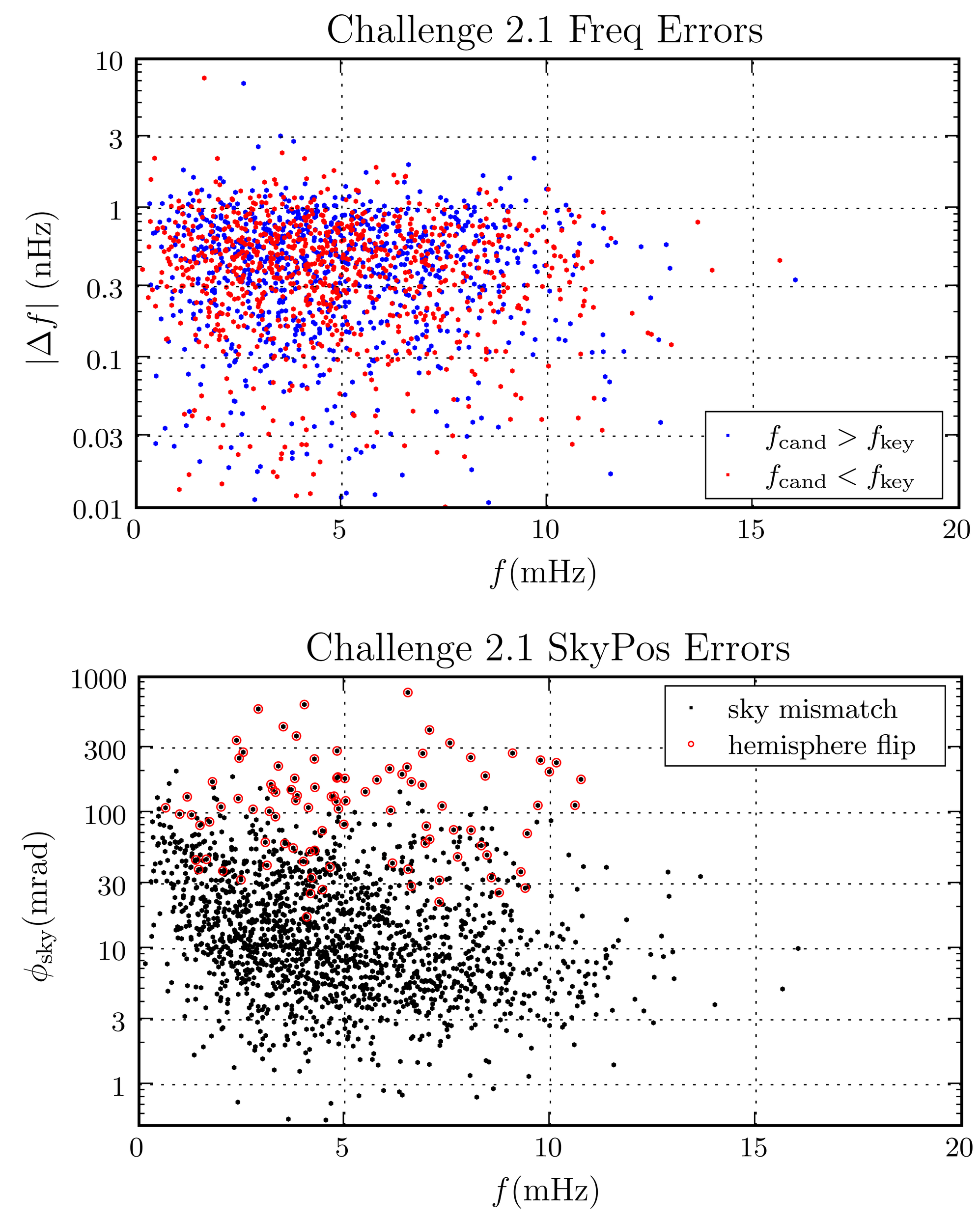

Figure 5: *Errors in estimated Doppler parameters: frequency Δf and skyposition $\phi_{\rm sky}$. Despite the use of the LWL the Doppler accuracy remains good. "Hemisphere flips" refer to signals that are found with the wrong sign of the latitude, i.e., $\beta_{\rm cand} \approx -\beta_{\rm key}$.*

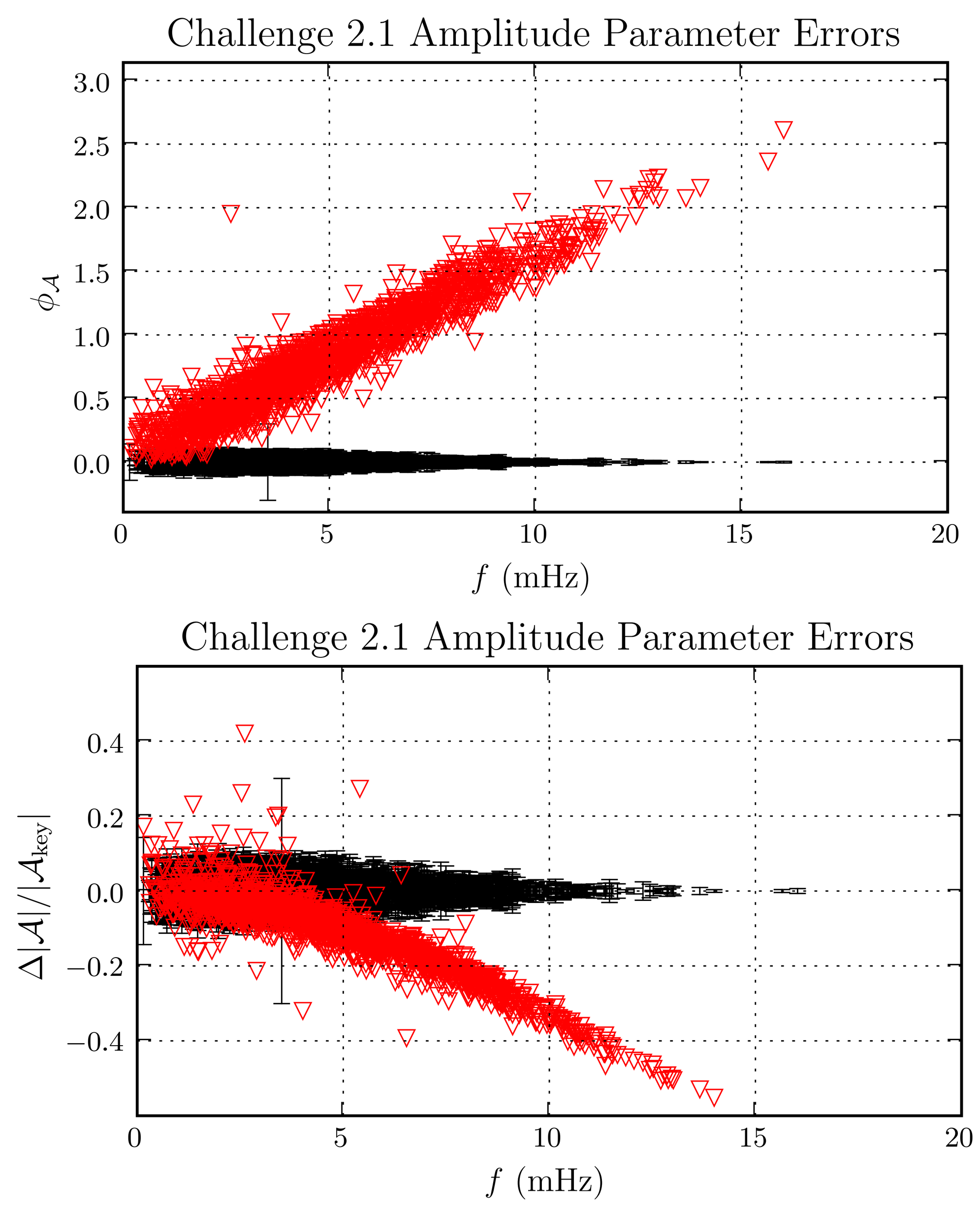

Figure 6: *Errors in estimated amplitude parameters $\mathcal{A}_{\rm cand}$: angle $\phi_{\mathcal{A}} \equiv \angle(\mathcal{A}_{\rm cand}, \mathcal{A}_{\rm key})$ and relative loss in magnitude $|\mathcal{A}_{\rm cand}|$ with respect to the injected $|\mathcal{A}_{\rm key}|$. These errors are strongly affected by the LWL, which only holds for $f \ll 10\,\mathrm{mHz}$. In each case, the black error bars show the expected one-sigma statistical error ($= 1/|\mathcal{A}_{\rm key}|$).*

Summary

Our MLDC2 pipeline is quite effective at separating primary candidates from secondary maxima: fewer than 1% of returned candidates are not attributable to bright sources ("false alarms" in table 1).

We are currently working to extend the $\mathcal{F}$-statistic code beyond the LWL by using the rigid adiabatic approximation. More work is also required to improve the detection efficiency of this pipeline, as only a fraction of the $\mathcal{O}(10,000)$ detectable sources were recovered.

References

[1] R. Prix and J. T. Whelan submitted to CQG; arXiv:0707.0128