Radial velocity data analysis with compressed sensing techniques

Nathan Hara, G. Boué, J. Laskar, A. Correia

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **464,** 1220–1246 (2017)
Advance Access publication 2016 September 8

doi:10.1093/mnras/stw2261

Radial velocity data analysis with compressed sensing techniques

Nathan C. Hara,[1]★ G. Boué,[1] J. Laskar[1] and A. C. M. Correia[1,2]

[1]*ASD/IMCCE, CNRS-UMR8028, Observatoire de Paris, PSL, UPMC, 77 Avenue Denfert-Rochereau, F-75014 Paris, France*
[2]*CIDMA, Departamento de Física, Universidade de Aveiro, Campus de Santiago, P-3810-193 Aveiro, Portugal*

Accepted 2016 September 7. Received 2016 September 5; in original form 2016 April 2

ABSTRACT

We present a novel approach for analysing radial velocity data that combines two features: all the planets are searched at once and the algorithm is fast. This is achieved by utilizing compressed sensing techniques, which are modified to be compatible with the Gaussian process framework. The resulting tool can be used like a Lomb–Scargle periodogram and has the same aspect but with much fewer peaks due to aliasing. The method is applied to five systems with published radial velocity data sets: HD 69830, HD 10180, 55 Cnc, GJ 876 and a simulated very active star. The results are fully compatible with previous analysis, though obtained more straightforwardly. We further show that 55 Cnc e and f could have been respectively detected and suspected in early measurements from the Lick Observatory and Hobby–Eberly Telescope available in 2004, and that frequencies due to dynamical interactions in GJ 876 can be seen.

Key words: methods: data analysis – techniques: radial velocities – planets and satellites: detection.

1 INTRODUCTION

1.1 Overview

Determining the content of radial velocity data is a challenging task. There might be several companions to the star, unpredictable instrumental effects as well as astrophysical jitter. Fitting separately the different features of the model might distort the residual and prevent one from finding small planets, as pointed out for instance by Anglada-Escudé, López-Morales & Chambers (2010) and Tuomi (2012). There might even be the cases where, due to aliasing and noise, the tallest peak of the periodogram is a spurious one while being statistically significant. To overcome those issues, recent approaches privilege the fitting of the whole model at once. In those cases, the usual framework is the maximization of an a posteriori probability distribution. In order to avoid being trapped in a sub-optimal solution, random searches such as Monte Carlo Markov chain methods or genetic algorithm are used (e.g. Gregory 2011; Ségransan et al. 2011). The goal of this paper is to suggest an alternative method using convex optimization, therefore offering a unique minimum and faster algorithms.

To do so, we will not try to find directly the orbital parameters of the planets but to unveil the true spectrum of the underlying continuous signal, which is equivalent. The power spectrum is often estimated with a Lomb–Scargle periodogram (Lomb 1976; Scargle 1982) or generalizations (Ferraz-Mello 1981; Cumming, Marcy & Butler 1999; Zechmeister & Kürster 2009). However, as said above, the estimation of the power spectrum with one frequency at a time has severe drawbacks. To improve the estimate, we introduce a priori information: the representation of exoplanetary signal in the Fourier domain is sparse. In other words, the number of sine functions needed to represent the signal is small compared to the number of observations. The Keplerian models are not the only ones to verify this assumption, stable planetary systems are quasi-periodic as well (e.g. Laskar 1993). By doing so, the periodogram can be efficiently cleaned (see Figs 1–5).

The field of signal processing devoted to the study of sparse signals is often referred to as 'compressed sensing' or 'compressive sampling' (Candès, Romberg & Tao 2006b; Donoho 2006) – though it is sometimes restricted to sampling strategies based on sparsity of the signal. The related methods show very good performances and are backed up by solid theoretical results. For instance, compressed sensing techniques allow one to recover exactly a spectrum while sampling it at a much lower rate than the Nyquist frequency (Mishali, Eldar & Tropp 2008; Tropp et al. 2010). Its use was advocated to improve the scientific data transmission in space-based astronomy (Bobin, Starck & Ottensamer 2008). Sparse recovery techniques are also used in image processing (e.g. Starck, Elad & Donoho 2005).

It seems relevant to add to that list a few techniques developed by astronomers to retrieve harmonics in a signal. In the next section, we show that even though the term 'sparsity' is not explicitly used (except in Bourguignon, Carfantan & Böhm 2007), some of the existing techniques have an equivalent in the compressed sensing literature. After those remarks on our framework, the paper is organized as follows: in Section 2, the theoretical background and the associated algorithms are presented. Section 3 presents in detail the procedure we developed for analysing radial velocity data. This one is applied in Section 4 to simulated observations and four real radial velocity data sets: HD 69830, HD 10180, 55 Cnc and GJ 876 and to a simulated very active star. The performance of

★ E-mail: nathan.hara@obspm.fr

the method is discussed in Section 5, and conclusions are drawn in Section 6.

1.2 Previous work

The goal of this paper is to devise a method to efficiently analyse radial velocity data. As it builds upon the retrieval of harmonics, the discussion will focus on spectral synthesis of unevenly sampled data (see Kay & Marple 1981; Schwarzenberg-Czerny 1998; Babu & Stoica 2010, for surveys).

First, let us consider the methods that are efficient to spot one harmonic at a time. The first statistical analysis is given by Schuster (1898). However, the statistical properties of Schuster's periodogram only hold when the measurements are equispaced in time. When this is not the case, one can use Lomb–Scargle periodogram (Lomb 1976; Scargle 1982) or its generalization consisting in adding a constant to the model (Ferraz-Mello 1981; Cumming et al. 1999; Reegen 2007; Zechmeister & Kürster 2009). More recently, Mortier et al. (2015) derived a Bayesian periodogram associated with the maximum of an a posteriori distribution. Also, Cumming (2004) and O'Toole et al. (2009) define the Keplerian periodogram, which measures the χ^2 of residuals after the fit of a Keplerian curve. One can remark that 'Keplerian' vectors defined by P, e, ω and M_0 form a family of vectors in which the sparsity of exoplanetary signals is enhanced.

These methods can be applied iteratively to retrieve several harmonics. In the context of radial velocity data processing, one searches for the peak of maximum power, then the corresponding signal is subtracted and the search is performed again. This procedure is very close to CLEAN (Roberts, Lehar & Dreher 1987), which relies on the same principle of maximum correlation and subtraction. One of the first general algorithm exploiting sparsity of a signal in a given set of vectors (Matching Pursuit, Mallat & Zhang 1993) relies on the same iterative process. This method was formerly known as forward stepwise regression (e.g. Bellmann 1975). To limit the effects of error propagation in the residuals, one can use the orthogonal matching pursuit algorithm (Pati, Rezaiifar & Krishnaprasad 1993; Tropp & Gilbert 2007). In that case, when a harmonic is found to have maximum correlation with the residuals, it is not directly subtracted. The next residual is computed as the original signal minus the fit of all the frequencies found so far. The CLEANest algorithm (Foster 1995) and frequency map analysis (Laskar 1988; Laskar, Froeschlé & Celletti 1992; Laskar 1993; Laskar 2003), though developed earlier, are particular cases of this algorithm. To analyse radial velocity data, Baluev (2009) and Anglada-Escudé & Tuomi (2012) introduce what they call respectively the 'residual periodogram' and the 'recursive periodogram', which can be seen as pushing that logic one step further. The principle is to re-fit at each trial frequency the previous Keplerian signals plus a sine at the considered frequency.

Besides the matching pursuit procedures, there are two other popular algorithms in the compressed sensing literature: convex relaxations (e.g. Tibshirani 1994; Chen, Donoho & Saunders 1998; Starck et al. 2005) and iteratively re-weighted least squares (IRWLS; e.g. Gorodnitsky & Rao 1997; Candès, Romberg & Tao 2006a; Donoho 2006; Daubechies et al. 2010). In the context of astronomy, Bourguignon et al. (2007) implement a convex relaxation method using ℓ_1 norm weighting (see equation 2) to find periodicity in unevenly sampled signals, and Babu et al. (2010) present an IRWLS algorithm named IAA to analyse radial velocity.

The methods presented above are apparently very different, yet they can be viewed as a way to bypass the brute force minimization of

$$\arg\min_{K,\omega,\phi} \sum_{i=1}^{m} \left(\boldsymbol{y}(t_i) - \sum_{j=1}^{k} K_j \cos(\omega_j t_i + \phi_j) \right)^2, \quad (1)$$

where $\boldsymbol{y}(\boldsymbol{t})$ is a vector made of m measurements, and $\boldsymbol{x}^\star = \arg\min f(\boldsymbol{x})$ denotes the element such that $f(\boldsymbol{x}^\star) = \min f(\boldsymbol{x})$ for a function f. This problem is very similar to 'best k-term approximation', and its link to compressed sensing has been studied in Cohen, Dahmen & Devore (2009) in the noise-free case. Solving that problem is suggested by Baluev (2013b) under the name of 'multifrequency periodograms'. However, finding that minimum by discretizing the values of $(K_j, \omega_j, \phi_j)_{j=1,\ldots,k}$ depends exponentially on the number of parameters, and the multifrequency periodograms could hardly handle more than three or four sines with conventional methods. However, with parallel programming on GPUs one can handle up to ≈25 frequencies depending on the number of measurements (Baluev 2013a). Jenkins et al. (2014) explicitly mention the above problem and suggest a tree-like algorithm to explore the frequency space. They analyse GJ 876 with their procedure and find six significant harmonics, which we confirm in Section 4.5.2.

Let us mention that searching for a few sources of periodicity in a signal is not always done with the Fourier space. When the shape of the repeating signal or the noise structure is not well known, other tests might be more robust. A large part of those methods consists in computing the autocorrelation function or folding the data at a certain period and look for correlation. See Engelbrecht (2013) for a survey or Zucker (2015, 2016) in the context of radial velocity measurements. Finally, we point out that the use of the sparsity of the signal is not specific to compressed sensing. The number of planets in a model is often selected via likelihood ratio tests. A model with an additional planet must yield a significant improvement of the evidence. In general, the model with $k+1$ planets $\mathcal{M}_{k+1}$ is selected over a model with k planet if $\Pr\{\boldsymbol{y}(\boldsymbol{t})|\mathcal{M}_{k+1}\}/\Pr\{\boldsymbol{y}(\boldsymbol{t})|\mathcal{M}_k\}$ is greater than 150 (see Tuomi et al. 2014), $\boldsymbol{y}(\boldsymbol{t})$ being the observations. Indeed, adding more parameters to the model automatically decreases the χ^2 of the residuals. Putting a minimum improvement of the χ^2 acts against overly complicated models.

The discussion above points that searching planets one after another is already in the compressed sensing paradigm: this iterative procedure is close to the orthogonal matching pursuit algorithm. Donoho, Elad & Temlyakov (2006) show that for a wide range of signals, this algorithm is outperformed by ℓ_1 relaxation methods. Does this claim still apply to radial velocity signals ? In this paper, this question is not treated in full generality, but we show the interest of ℓ_1 relaxation on several examples. To address that question more directly, it is shown in Appendix C that in some cases, the tallest peak of the periodogram is spurious but ℓ_1 minimization prevents one from being misled.

2 METHODS

2.1 Minimization problem

Techniques based on sparsity are thought to enforce the 'Occam's razor' principle: the simplest explanation is the best. To apply that principle, we must have an idea of 'how' the signal is simple. In the compressed sensing framework (or compressive sampling), this is done by selecting a set of vectors $\mathcal{A} = (\boldsymbol{a}_j(\boldsymbol{t}))_{j\in I}$ such that the signal to be analysed $\boldsymbol{y}(\boldsymbol{t})$ is represented by a linear combination of

a few elements of $\mathcal{A}$. Such a set is often called the 'dictionary' and can be finite or not (the set of indices I can be finite or infinite). It is here made of vectors $\boldsymbol{a}_{-\omega}(\boldsymbol{t}) = \mathrm{e}^{-\mathrm{i}\omega t}$ and $\boldsymbol{a}_{\omega}(\boldsymbol{t}) = \mathrm{e}^{\mathrm{i}\omega t}$, where $\boldsymbol{t}$ is the array of measurement times.

Before going into the details, let us define some quantities.

(i) $\boldsymbol{y}(\boldsymbol{t})$ denotes the vector of observations at times $t = t_1, \ldots, t_m$, $\boldsymbol{y}(\boldsymbol{t}) \in \mathbb{R}^m$ for radial velocity data sets.

(ii) The ℓ_p norm of a complex or real vector x with n components is defined as

$$\|\boldsymbol{x}\|_{\ell_p} := \left(\sum_{k=1}^{n} |x_k|^p \right)^{1/p} \tag{2}$$

for $p > 0$. In particular, $\|\boldsymbol{x}\|_{\ell_1}$ is the sum of absolute values of the vector components and $\|\boldsymbol{x}\|_{\ell_2} = \sqrt{\sum_{k=1}^{n} |\boldsymbol{x}_k|^2}$ is the usual Euclidian norm. When $p = 0$, $\|\boldsymbol{x}\|_{\ell_0}$ is the number of non-zero components of $\boldsymbol{x}$.

(iii) For a function f defined on a set E, $\arg\min_{\boldsymbol{x}\in E} f(\boldsymbol{x})$ is the element for which the minimum is attained, that is $\boldsymbol{x}^\star$ of E such that $f(\boldsymbol{x}^\star) = \min_{\boldsymbol{x}\in E} f(\boldsymbol{x})$. We denote by the superscript $\star$ the solution of the minimization problem under consideration. In all the cases considered here except equations (1) and (3), the minimum is attained as we consider convex functions on convex sets.

Let us consider combinations of S elements of the dictionary $(\boldsymbol{a}_j(\boldsymbol{t}))_{j=1,\ldots,S}$ and their corresponding amplitudes x_j. To enhance the sparsity of the representation, one can think of solving

$$\underset{\substack{\boldsymbol{a}_{j(t)}\in\mathcal{A}\\ S\in\mathbb{C}}}{\arg\min} \quad S \quad \text{s.t.} \quad \left\| \sum_{j=1}^{S} x_j \boldsymbol{a}_j(\boldsymbol{t}) - \boldsymbol{y}(\boldsymbol{t}) \right\|_{\ell_2} \leqslant \epsilon \tag{3}$$

that is finding the smallest number of elements of $\mathcal{A}$ required to approximate $\boldsymbol{y}(\boldsymbol{t})$ with a certain tolerance ϵ. This one is a priori a combinatorial problem which seems unsolvable if $\mathcal{A}$ is infinite or of an exponential complexity if the dictionary is finite. In the latter case, $\mathcal{A}$ can be viewed as an $m \times n$ matrix $\mathbf{A}$. In that case, on can re-write equation (3) like

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}\in\mathbb{C}^n} \|\boldsymbol{x}\|_{\ell_0} \quad \text{s.t.} \quad \|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}(\boldsymbol{t})\|_{\ell_2} \leqslant \epsilon. \tag{4}$$

This problem is in general combinatorial (Ge, Jiang & Ye 2011), therefore computationally intractable. Fortunately, when replacing the ℓ_0 norm by the ℓ_1 norm,

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}\in\mathbb{C}^n} \|\boldsymbol{x}\|_{\ell_1} \quad \text{s.t.} \quad \|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}(\boldsymbol{t})\|_{\ell_2} \leqslant \epsilon, \tag{5}$$

the problem becomes convex and still enhances sparsity efficiently. In the signal processing literature, this problem is referred to as basis pursuit denoising (Chen et al. 1998), and is sometimes denoted by BP_ϵ. At this point, one can ask what is lost by considering equation (5) instead of equation (3). Let us cite a few results – among many: when $\boldsymbol{y}(\boldsymbol{t})$ is noise free, Donoho (2006) shows that under certain hypotheses the solution to equation (5) is equal to the solution of equation (3); more generally, denoting by $\boldsymbol{y}_t = \mathbf{A}\boldsymbol{x}_t$ the true signal, such that $\boldsymbol{y} = \boldsymbol{y}_t + \boldsymbol{e}$, $\boldsymbol{e}$ being the error, there is a theoretical bound on $\|\mathbf{A}\boldsymbol{x}^\star - \boldsymbol{y}_t\|_{\ell_2}$ (Candès et al. 2006b). One can also obtain constraints on $\|\boldsymbol{x}^\star - \boldsymbol{x}_t\|_{\ell_2}$ or conditions to have $\mathrm{supp}(\boldsymbol{x}^\star) \subset \mathrm{supp}(\boldsymbol{x}_t)$, where $\mathrm{supp}(\boldsymbol{x})$ is the set of indices with x being non-zero (e.g. Donoho et al. 2006). In summary, there are results guaranteeing the performance for denoising, compression and also for inverse problems, the search for planets being a particular case of the latter.

These results apply to a finite dictionary $\mathcal{A}$, but the periods of the planets could be anywhere: $\mathcal{A}$ is infinite for our purposes. We will eventually go back to solving a modified version of the discrete problem (5) and smooth its solution with a moving average. Beforehand, we will present in the next section what seems to be the most relevant theoretical background for our studies, 'atomic norm minimization', in particular used in 'super-resolution theory' (Candès & Fernandez-Granda 2014). This one will give guidelines to improve our procedure.

2.2 Atomic norm minimization

If $\mathcal{A}$ is infinite, the ℓ_1 norm cannot be used straightforwardly. Chandrasekaran et al. (2010) suggest to use an 'atomic norm' that extends equation (5) to infinite dictionaries. Practical methods to solve the new minimization problem are designed in Candès & Fernandez-Granda (2013) and Tang et al. (2013b). The atomic norm $\|\boldsymbol{y}\|_{\mathcal{A}}$, of $\boldsymbol{y} \in \mathbb{R}^m$ or $\mathbb{C}^m$ defined for a dictionary $\mathcal{A}$, is the smallest ℓ_1 norm of a combination of vectors of the dictionary reproducing y:

$$\|\boldsymbol{y}\|_{\mathcal{A}} = \inf \left\{ \sum_j |x_j|, \boldsymbol{y} = \sum_j x_j \boldsymbol{a}_j(\boldsymbol{t}) \right\}. \tag{6}$$

If the observations were not noisy, computing the atomic norm of $\boldsymbol{y}$ would be sufficient. As this is obviously not the case, the following problem is considered,

$$\boldsymbol{u}^\star = \underset{\boldsymbol{u}\in\mathbb{C}^m}{\arg\min} \quad \|\boldsymbol{u} - \boldsymbol{y}(\boldsymbol{t})\|_{\ell_2}^2 + \lambda \|\boldsymbol{u}\|_{\mathcal{A}}, \tag{7}$$

where λ is a positive real number fixed according to the noise. This problem is often referred to as atomic norm denoising. The coefficient λ can be interpreted as a Lagrange multiplier, and this problem can be seen as maximizing a posterior likelihood with a prior on $\boldsymbol{u}$. The quantities we are interested in are the dictionary elements $\boldsymbol{a}_j^\star$ and the coefficients $\boldsymbol{x}^\star$ selected by the minimization, where $\boldsymbol{u}^\star = \sum_{j=1}^{S^\star} x_j^\star \boldsymbol{a}_j^\star(\boldsymbol{t})$.

2.3 More complex noise models

If exoplanetary signals are arguably a sum of sines plus noise, the noise variance is not constant. Even more, the noise might not be independent nor Gaussian. Recent papers such as Tuomi et al. (2013) and Rajpaul et al. (2015) stress that the detection efficiency and robustness improve as the noise model becomes more realistic. Aigrain et al. (2011) suggest to consider the RV time series as Gaussian processes: the noise $\boldsymbol{n}(\boldsymbol{t})$ is then characterized by its covariance matrix $\mathbf{V}$ which is such that $V_{kl} = \mathbb{E}\{\boldsymbol{n}(t_k)\boldsymbol{n}(t_l)\}$, $\mathbb{E}$ being the mathematical expectancy. When the noise is stationary, by definition there exists a covariance function R such that $V_{kl} = R(|t_l - t_k|)$; therefore, choosing $\mathbf{V}$ is equivalent to choosing R. This approach is similar to Sulis, Mary & Bigot (2016), which normalizes the periodogram by the power spectrum of the stationary part of the stellar noise. The similarity comes from the fact that the power spectrum of the noise is $P(\omega) = |\mathcal{F}(R)|^2$, where $\mathcal{F}$ denotes the Fourier transform.

Here, the noise is assumed to be Gaussian of covariance matrix $\mathbf{V}$. In that case, the logarithm of the likelihood is (e.g. Baluev 2011, equation 21; Pelat 2013)

$$\ln(L) = -\frac{m}{2}\ln(2\pi) - \frac{1}{2}\det(\mathbf{V}) - \frac{1}{2}(\boldsymbol{y} - \mathbf{A}\boldsymbol{x})^T \mathbf{V}^{-1} (\boldsymbol{y} - \mathbf{A}\boldsymbol{x}), \tag{8}$$

where the subscript T denotes the matrix transpose. Assuming that the matrix $\mathbf{V}$ is fixed, we wish to minimize $(\boldsymbol{y}-\mathbf{A}\boldsymbol{x})^T\mathbf{V}^{-1}(\boldsymbol{y}-\mathbf{A}\boldsymbol{x})$. If $\mathbf{V}^{-1}$ admits a square root, then $\mathbf{W}$ is chosen such that $\mathbf{W}^2=\mathbf{V}^{-1}$. This is the case when $\mathbf{V}$ is symmetric positive definite, which is the case for covariance matrices of stationary processes. Consequently, $\|\mathbf{W}(\mathbf{A}\boldsymbol{x}-\boldsymbol{y})\|_{\ell_2}^2=(\boldsymbol{y}-\mathbf{A}\boldsymbol{x})^T\mathbf{V}^{-1}(\boldsymbol{y}-\mathbf{A}\boldsymbol{x})$ is always ensured for Gaussian noises. We then obtain the minimization

$$\underset{\boldsymbol{u}\in\mathbb{C}^m}{\arg\min}\quad \|\mathbf{W}(\boldsymbol{u}-\boldsymbol{y}(\boldsymbol{t}))\|_{\ell_2}^2+\lambda\|\boldsymbol{u}\|_{\mathcal{A}}. \tag{9}$$

Handling problem (5) with correlated measurements and noise has been investigated by Arildsen & Larsen (2014). However, to the best of our knowledge, the formulation above is not mentioned in the literature; thus, we will briefly discuss its features.

The ability of problem (5) to unveil the true non-zero coefficients of x improves as the so-called mutual coherence of matrix $\mathbf{A}$ diminishes (Donoho et al. 2006). This one is defined as the maximum correlation between two column vectors of $\mathbf{A}$. We here consider a weight matrix, but we can go back to the previous problem by noting that $\mathbf{W}(\mathbf{A}\boldsymbol{x}-\boldsymbol{y})$ can be re-written as $\mathbf{A}'\boldsymbol{x}-\boldsymbol{y}'$, where $\mathbf{A}'=\mathbf{W}\mathbf{A}$ and $\boldsymbol{y}'=\mathbf{W}\boldsymbol{y}$. If we now consider two column vectors of $\mathbf{A}'$, $\boldsymbol{a}_1'=\mathbf{W}\boldsymbol{a}_1$ and $\boldsymbol{a}_2'=\mathbf{W}\boldsymbol{a}_2$, their correlation is $\boldsymbol{a}_1'^T\boldsymbol{a}_2'=\boldsymbol{a}_1W^TW\boldsymbol{a}_2=\boldsymbol{a}_1V^{-1}\boldsymbol{a}_2$. In other words, introducing a matrix $\mathbf{W}$ only comes down to changing the scalar product. This should not be surprising. The matched filter technique (Kay 1993) proposes to detect a model $\boldsymbol{x}$ in a signal $\boldsymbol{s}=\boldsymbol{x}+\boldsymbol{n}$, where $\boldsymbol{n}$ is a noise of covariance matrix $\mathbf{V}$ if $\boldsymbol{x}\mathbf{V}^{-1}\boldsymbol{s}\leq\gamma$, where γ is a threshold. This means if the correlation is sufficient for a non-trivial scalar product.

In the case of an independent Gaussian noise, its covariance matrix $\mathbf{V}$ is diagonal and its elements are σ_k^2, where σ_k is the measurement error at time t_k. $\mathbf{W}$ is defined as $\mathbf{V}^{-1/2}$ so is a diagonal matrix of elements $w_{kk}=1/\sigma_k$. Therefore, $\boldsymbol{a}_1'^T\boldsymbol{a}_2'=\boldsymbol{a}_1\mathbf{W}^T\mathbf{W}\boldsymbol{a}_2=\sum_{k=1}^n\frac{\boldsymbol{a}_1(t_k)\boldsymbol{a}_2(t_k)}{\sigma_k^2}$. This is compatible with the behaviour we intuitively expect: the less precise is the measurement, the lesser the correlation between the signals matter through the weighting by σ_k.

Unfortunately, having a non-identically independent distributed (i.i.d.) Gaussian noise model biases the estimates of the true signals as it acts as a frequency filter. Whether this bias prevents one from having the benefits of a correct noise model is discussed in Appendix B. We show that choosing an appropriate weight matrix $\mathbf{W}$ indeed allows us to see signals that would be buried in the red noise otherwise.

3 IMPLEMENTATION

3.1 Overview

As said above, stable planetary systems are quasi-periodic. This means in particular that radial velocity measurements are well approximated by a linear combination of a few vectors $\mathrm{e}^{-\mathrm{i}\omega t}$ and $\mathrm{e}^{\mathrm{i}\omega t}$. The minimization problem (7) seems therefore well suited for searching for exoplanets. This section is concerned with the numerical resolution, and the numerous issues it raises: the numerical scheme to be used, the choice of the algorithm parameters and the evaluation of the confidence in a detection.

Solving equation (7) is done either by reformulating it as a quadratic program (Candès & Fernandez-Granda 2013; Chen & Chi 2014; Tang et al. 2013b) or by discretizing the dictionary (Tang, Bhaskar & Recht 2013a). The first one necessitates to see the sampling as a regularly spaced one with missing samples. As the measurement times are far from being equispaced in the considered applications, the required time discretization results in large matrices. Therefore, the second approach is used.

Let us pick a set of frequencies equispaced with interval $\Delta\omega$, $\Omega=\{\omega_k=k\Delta\omega,\ k=0,\ldots,n\}$ and an $m\times 2n$ matrix $\mathbf{A}$ whose columns are $\mathrm{e}^{-\mathrm{i}\omega_k t}$ and $\mathrm{e}^{\mathrm{i}\omega_k t}$. In that case, equation (9) reduces to

$$\underset{\boldsymbol{x}\in\mathbb{C}^{2n}}{\arg\min}\quad \|\mathbf{W}(\mathbf{A}\boldsymbol{x}-\boldsymbol{y})\|_{\ell_2}^2+\lambda\|\boldsymbol{x}\|_{\ell_1}, \tag{10}$$

which is often referred to as the Least Absolute Shrinkage and Selection Operator (LASSO) problem when W is the identity matrix. As the parameter λ is not so easy to tune, an equivalent formulation of discretized equation (9) is chosen,

$$\boldsymbol{x}^\star=\underset{\boldsymbol{x}\in\mathbb{C}^{2n}}{\arg\min}\quad \|\boldsymbol{x}\|_{\ell_1}\quad \text{s. t.}$$
$$\|\mathbf{W}(\mathbf{A}\boldsymbol{x}-\boldsymbol{y})\|_{\ell_2}\leqslant\epsilon, \tag{11, BP$_{\epsilon,\mathbf{W}}$}$$

where ϵ is a positive number. By 'equivalent', we mean there exists a λ_ϵ such that the solution of equation (10) is equal to the solution of equation (11,BP$_{\epsilon,W}$) (Rockafellar 1970). As this problem will often be referred to, we add to the equation number BP$_{\epsilon,W}$ in the rest of the text, BP standing for basis pursuit. There are several codes written to solve equation (5). The existing codes we have tested for analysing radial velocity data sets are ℓ_1-magic (Candès et al. 2006a), SparseLab (Donoho 2006), NESTA (Becker, Bobin & Candès 2011), CVX (Grant & Boyd 2008), spectral compressive sampling (Duarte & Baraniuk 2013) and SPGL1 (van den Berg & Friedlander 2008). The latter gave the best results in general for exoplanetary data and consequently is the one we selected (the code can be downloaded from this link[1]).

The solution of equation (11,BP$_{\epsilon,W}$) offers an estimate for the periods, but the efficiency of the method can be improved by using a moving average on $\boldsymbol{x}^\star$, to approximate better equation (9). Indeed, if a sine of frequency ω_0 and amplitude K is in the signal, corollary 1 (Tang et al. 2013a) shows that the solution of equation (5) $\boldsymbol{x}^\star$ verifies

$$K\quad\approx\sum_{|\omega_k|\in[\omega_0-\eta,\omega_0+\eta]}x^\star(\omega_k) \tag{12}$$

rather than $|x(\omega_0)|\approx K$. The coefficients $x^\star(\omega_k)$ are added up for ω_k lying in a certain interval of length 2η (see Section 3.6).

Finally, the confidence in the detection must be estimated. Problem (11,BP$_{\epsilon,W}$) selects significant frequencies in the data, but the estimates of their amplitude are biased due to the ℓ_1 norm minimization. To obtain unbiased amplitudes, we first check that the peaks are not aliases of each other. Then the most significant peaks are fitted until non-significant residuals are obtained (see Section 3.7.4).

In summary, the method follows a seven-step process.

(i) Pre-process the data: remove the mean in radial velocity data or an estimate of the stellar noise.

(ii) Choose the discrete grid Ω, tolerance ϵ, weighting matrix $\mathbf{W}$ and the width η of the interval over which the result of equation (11,BP$_{\epsilon,W}$) is averaged.

(iii) Define the dictionary $\mathcal{A}$ and normalize the columns of $\mathbf{WA}$.

(iv) Run the program solving the convex optimization (11,BP$_{\epsilon,W}$) to obtain $\boldsymbol{x}^\star$.

(v) Denoting $\Omega=[\omega_{\min},\omega_{\max}]$ for each frequency $\omega\in\{\omega_{\min}+\eta,\ldots,\omega_{\max}-\eta\}$, sum up the amplitudes of $x^\star(\omega')$ from $\omega'\in[-\omega-\eta,-\omega+\eta]\cup[\omega-\eta,\omega+\eta]$ to obtain a smoothed figure $\boldsymbol{x}^\sharp$.

[1] https://www.math.ucdavis.edu/~mpf/spgl1/supplement.html

(vi) Plot $\boldsymbol{x}^\sharp$ as a function of the frequencies or the periods.
(vii) Evaluate the significance of the main peaks (Fig. 6).

Each of these steps is detailed in the following sections.

3.2 Optimization routine

Many solvers can handle problem (5); however, their precision and speed vary. Among the solvers tested, SPGL1 (van den Berg & Friedlander 2008) gives the best results in general. This one has several user-defined parameters such as a stopping criterion that must be tuned. For a given tolerance, this one is $\frac{|\|\mathbf{A}\boldsymbol{x}-\boldsymbol{y}\|_{\ell_2}-\epsilon|}{\max(1,\|\mathbf{A}\boldsymbol{x}-\boldsymbol{y}\|_{\ell_2})} < \text{tol}$. The default parameters seem acceptable, in particular $\text{tol} = 10^{-4}$.

3.3 Dictionary A

To estimate the spectrum, a natural choice for the columns of matrix $\mathbf{A}$ is $(\mathrm{e}^{-\mathrm{i}\omega \boldsymbol{t}}, \mathrm{e}^{\mathrm{i}\omega \boldsymbol{t}})$. However, the data might not contain only planetary signals. In the case of a binary star, a linear trend t and a quadratic term t^2 are added. If the star is active, the ancillary measurements are also added.

The method described in Section 3 is applicable to a wider range of dictionary. As the timespan of the observations is in general a few years, the signal might be more sparsely represented either by Poisson terms $[(a_0 + a_1 t + a_2 t^2 + \cdots)\cos(\omega t + \phi)]$ or Keplerian motions. In the latter case, column vectors would be of the form $\frac{r}{a}\mathrm{e}^{\mathrm{i}\boldsymbol{\nu}(\boldsymbol{t})}$, where $\boldsymbol{\nu}(\boldsymbol{t})$ is a vector of true anomalies depending on the period P, eccentricity e and initial mean anomaly M_0 (or any combination of three variables that cover all possible orbits). Unfortunately, the size of A increases exponentially with the number of parameters describing the dictionary elements (here P, e, M_0).

3.4 Pre-processing

Theoretical results in Tang et al. (2013a) guarantee that the solution to equation (5) will be close to equation (7) as the discretization gets finer, provided the dictionary is continuous. As linear trends or stellar activity-related signals are not sine, removing these from the data before solving equation (11,$\mathrm{BP}_{\epsilon,W}$) is crucial. The mean, a linear trend and estimates of the stellar noise can be fitted and removed. We reckon this is contrary to the philosophy of fitting the whole model at once. However, the vectors fitted are included again in the dictionary which allows us to mitigate the distortions induced by their removal.

Secondly, to make the precision of the SPGL1 solver independent from the value of $\mathbf{W}\boldsymbol{y}$, the weighted observations $\mathbf{W}\boldsymbol{y}$ are normed by $\|\mathbf{W}\boldsymbol{y}\|_{\ell_2}$, and the columns of the matrix $\mathbf{WA}$ are also normed. Denoting by $\boldsymbol{y}' = \frac{1}{\epsilon}\mathbf{W}\boldsymbol{y}/\|\mathbf{W}\boldsymbol{y}\|_{\ell_2}$ and $\mathbf{A}' = \frac{1}{\epsilon}(\mathbf{W}A_k/\|\mathbf{W}A_k\|_{\ell_2})_{k=1,\ldots,n}$, we set in input of the solver

$$\arg\min_{\boldsymbol{x}\in\mathbb{C}^n} \|\boldsymbol{x}\|_{\ell_1} \quad \text{s.t.} \quad \left\|\mathbf{A}'\boldsymbol{x} - \boldsymbol{y}'\right\|_{\ell_2} \leqslant 1, \tag{13}$$

to always be in the same kind of use of the solver and ensure that the accuracy of the result does not depend on its units. Going back to the correct units in the post-processing step is described in Section 3.6.

3.5 Tuning

Choice of $\mathbf{W}$. We have seen in Section 2.3 that the weight matrix $\mathbf{W}$ is characterized by the covariance function R via $\mathbf{W}_{kl} = R(|t_k - t_l|)$. Several forms for the covariance functions were suggested (e.g. Rajpaul et al. 2015). Here we only consider exponential covariances, which are

$$\begin{aligned} R(\Delta t) &= \sigma_R^2 \mathrm{e}^{-\frac{|\Delta t|}{\tau}}, \quad \Delta t \neq 0 \\ R(0) &= \sigma_W^2 + \sigma_R^2, \end{aligned} \tag{14}$$

where the subscripts W and R stand respectively for white and red. As the red and white noises are here supposed independent, the covariance function of their sum is the sum of their covariance functions. Therefore, the matrix $\mathbf{W}$ is such that its diagonal terms are $V_{kk} = \sigma_k^2 + \sigma_W^2 + \sigma_R^2$ and $V_{kl} = \sigma_R^2 \mathrm{e}^{-\frac{|t_k - t_l|}{\tau}}$ for $k \neq l$.

Choice of Ω. We have two parameters to choose: the grid span and the grid spacing. For the first one, we take 1.5 cycles d^{-1} as a default value but it is also advisable to re-do the analysis for 0.95 cycles d^{-1}, as discussed in the examples in Section 2. We ensure that if the signal is made of sinusoids (a.k.a. it is quasi-periodic), there exists at least one vector x verifying $\|\mathbf{W}(\mathbf{A}\boldsymbol{x} - \boldsymbol{y})\|_{\ell_2} < \epsilon$ that has the correct ℓ_0 norm. Let us consider a signal made of p pure sinusoids sampled at times $\boldsymbol{t} = (\boldsymbol{t}_k)_{k=1,\ldots,m}$, $\boldsymbol{y}(\boldsymbol{t}) = \sum_{j=1}^{p} c_j \mathrm{e}^{\mathrm{i}\omega_j \boldsymbol{t}}$. Assuming that the frequencies on the grid are regularly spaced with step $\Delta\omega$, this leads to the condition (see Appendix A for calculation details)

$$\Delta\omega \leqslant \frac{4}{T} \arcsin \frac{\epsilon}{2\sqrt{\sum_{j=1}^{p} |c_j|^2}\sqrt{\sum_{k=1}^{m} \frac{1}{\sigma_k^2}}}. \tag{15}$$

Let us note that the values of c_j are a priori unknown, so the term $\sqrt{\sum_{j=1}^{p} |c_j|^2}$ has to be approximated. Supposing the signal is made of sinusoids plus small noise, $\sqrt{\sum_{j=1}^{p} |c_j|^2} \approx \|\boldsymbol{y}\|_{\ell_2}/\sqrt{m}$. Furthermore, it must be ensured that all possible significant frequencies are in the signal.

The choice of the grid spacing can be based on other criteria: Stoica & Babu (2012) suggest to choose a spacing such that the 'practical rank of matrix $M_{kl} = \mathrm{e}^{\mathrm{i}\Delta\omega(t_k - t_l)}$ is equal to one. This term designates the number of singular values above a certain threshold. Here the condition states that only one singular value is non-negligible. Let us also mention that one can perform the reconstruction with different grids and average out the results. However, this approach does not practically generate better results than using a finer grid.

Choice of ϵ. The error is due to two sources: grid discretization which gives an error ϵ_{grid} and noise, which yields ϵ_{noise}. Supposing the noise is Gaussian, denoting by $\boldsymbol{y}_t$ the underlying non-noisy observations, $\|\mathbf{W}(\boldsymbol{y}_t - \boldsymbol{y})\|_{\ell_2}^2$ as a function of random variable $\boldsymbol{y} = \boldsymbol{y}_t + \boldsymbol{n}$ follows a χ^2 distribution with m degrees of freedom. Denoting its cumulative distribution function (CDF) by $F_{\chi^2_m}$, the probability $1 - \alpha$ that the true signal $\boldsymbol{y}_t$ is in the set $\{\boldsymbol{y}', \|\mathbf{W}(\boldsymbol{y}' - \boldsymbol{y})\|_{\ell_2}^2 \leq \epsilon_{\text{noise}}\}$ is

$$F_{\chi^2_m}(\epsilon_{\text{noise}}^2) = 1 - \alpha. \tag{16}$$

The bound ϵ_{noise} is determined according to the equation above for a small α. Once ϵ_{noise} is chosen, rearranging equation (15) gives a minimal value of ϵ_{grid} that ensures that a signal with a correct ℓ_0 norm exists,

$$\epsilon_{\text{grid}} = 2\sqrt{\sum_{j=1}^{p} |c_j|^2}\sqrt{\sum_{k=1}^{m} \frac{1}{\sigma_k^2}} \sin \frac{\Delta\omega T_{\text{obs}}}{4}. \tag{17}$$

An alternative is to set ϵ to zero and let the algorithm find a representation for the noise, which will not be sparse. In that case,

one must obviously not perform the re-normalization of the columns of $\mathbf{WA}$ by ϵ of Section 3.4. Below a certain amplitude, a 'forest' of peaks would be seen on the ℓ_1-periodogram. This has the advantage to give an estimation of the noise structure. However, this method is more sensitive to the solver inner uncertainties and requires more time; it was not retained for this work.

Choice of η. See the next section.

3.6 Post-processing

Once the solution to equation (11,$\mathrm{BP}_{\epsilon,W}$) is computed, the spectrum $\boldsymbol{x}^\star$ is filtered with a moving average. We expect from discretization (9) that the frequencies might leak to close frequencies. Indeed, the amplitude of the solution to equation (11,$\mathrm{BP}_{\epsilon,W}$) might be untrustworthy. When the signal is made of several frequencies, the solution might overestimate the one with the greatest amplitude, and underestimate the others; this problem arises especially when less than a hundred observations are available. To mitigate this effect, one can sum up the contribution of subsequent frequencies and estimate the amplitude of the resulting signal. If $\boldsymbol{x}^\star$ is the solution to equation (11,$\mathrm{BP}_{\epsilon,W}$), denoting by $x^\star(\omega)$ the coefficient corresponding to frequency ω, we compute

$$\hat{\boldsymbol{y}}_\omega(\boldsymbol{t}) = \|\mathbf{W}\boldsymbol{y}\|_{\ell_2} \sum_{\omega' \in \Omega} \frac{x^\star(\omega')\boldsymbol{a}_{\omega'}(\boldsymbol{t})}{\|\mathbf{W}\boldsymbol{a}_{\omega'}(\boldsymbol{t})\|_{\ell_2}}, \quad \omega - \eta \leqslant |\omega'| \leqslant \omega + \eta \tag{18}$$

where $\boldsymbol{a}_{\omega'}(\boldsymbol{t})$ is the column of A corresponding to frequency ω'. The terms $\|\mathbf{W}\boldsymbol{y}\|_{\ell_2}$ and $1/\|\mathbf{W}\boldsymbol{a}_{\omega'}(\boldsymbol{t})\|_{\ell_2}$ appear because the columns of WA and the weighted observations Wy were normalized in Section 3.4. The vector $\hat{\boldsymbol{y}}_\omega(\boldsymbol{t})$, $t = t_1, \ldots, t_m$ is approximately a sine function; the new estimation of the signal power is

$$x^\sharp(\omega) = \max_{t_1..t_m} |\hat{y}_\omega(t_k)|. \tag{19}$$

Other estimates are possible, such as the power of a sine at frequency ω fitted on $\hat{y}(\omega)$. Though the choice of η is heuristic, corollary 1 of Tang et al. (2013a) is used as a guideline. It indeed states that the summed amplitudes of coefficients of $\boldsymbol{x}^\star$ within a certain distance η_0 from the actual peak in the signal tend to the appropriate value as the discretization step tends to zero. In the proof, they choose ϵ such that the balls of width η_0 centred around the true peaks have a null intersection. Thus, it seems reasonable to select η as the largest interval within which the probability to distinguish frequencies is low. Values such as $\approx 0.5\pi/T_{\mathrm{obs}}$ to π/T_{obs} are robust in practice.

3.7 Significance and uncertainties

3.7.1 Detection threshold

It is simple to associate a 'global' false alarm probability (FAP) to the ℓ_1-periodogram similar to the classical FAP of the Lomb–Scargle periodogram (Scargle 1982, equation 14). Let us consider the probability that '$x = 0$ is not a solution knowing the signal is pure independent Gaussian noise'. Denoting this probability $\tilde{\alpha}$, following notations of Section 2.1, $\epsilon^2 = {F_{\chi^2}}^{-1}(1-\tilde{\alpha})$. As $\epsilon \approx \epsilon_{\mathrm{noise}}$, the value of $\tilde{\alpha}$ is close to the user-defined parameter α. In the Lomb–Scargle case, the FAP obeys: 'if the maximum of the periodogram is z, then the FAP is $\beta(z)$', where β is an increasing function of z [often taken as $\beta(z) = 1 - (1 - \mathrm{e}^{-z})^M$, where M is a parameter fitted with numerical simulations; Scargle 1982; Horne & Baliunas 1986; Cumming 2004. Here the formulation is 'if the solution to equation (11,$\mathrm{BP}_{\epsilon,W}$) is not zero, then a signal has been detected with an FAP lower than or equal to α'.

3.7.2 Statistical significance of a peak

The discussion above points out similarities with the FAP defined for periodograms. This one and the global FAP share in particular that they only allow one to reject the hypothesis that the signal is pure Gaussian noise of covariance matrix $\mathbf{W}$. However, the problem is rather to determine if a given peak indicates a true underlying periodicity, and if this one is due to a planet.

In that scope, our goal is to test if the harmonics spotted by the ℓ_1-periodogram are statistically significant. Ultimately, one can use statistical hypothesis testing, which can be time consuming. To quickly assess the significance of the peaks, two methods seem to be efficient.

(i) Re-sampling: taking off randomly 10–20 per cent of the data and re-computing the ℓ_1-periodogram. The peaks that show great variability should not be trusted.

(ii) Using the formulae of the 'residual/recursive periodograms' (Cumming 2004; Baluev 2008, 2009, 2015a; Anglada-Escudé & Tuomi 2012).

The first case is easy to code and has the advantage to implement implicitly a time–frequency analysis. Indeed, we might expect from stellar variability some wavelet-like contributions: a signal with a certain frequency arises and then vanishes. The timespan of observation might be short enough so that feature is mistaken for a truly sinusoidal component. By taking off some of the measurements, we can see if the amplitude of a given frequency varies through time. However, this method requires to re-compute the ℓ_1-periodogram several times and might not be suited for systems with numerous measurements.

3.7.3 Model

As the re-sampling approach is straightforward to code, we will now focus on the recursive periodogram formulae. These ones should be useful for readers more interested in speed than comprehensiveness. In this section, the relevant signal models are defined. We consider that the signal is of the form

$$f_K\left(\boldsymbol{\theta}_0, \left(\boldsymbol{\theta}_{K\,j}\right)_{j=1..np}\right) = \text{non-planetary}\,(\boldsymbol{\theta}_0) + \sum_{j=1}^{n_p} \text{Keplerian}_j\left(\boldsymbol{\theta}_{K\,j}\right) \tag{20}$$

or

$$f_C\left(\boldsymbol{\theta}_0, \left(\boldsymbol{\theta}_{C\,j}\right)_{j=1..np}\right) = \text{non-planetary}\,(\boldsymbol{\theta}_0) + \sum_{j=1}^{n_p} \text{Circular}_j\left(\boldsymbol{\theta}_{C\,j}\right). \tag{21}$$

That is a sum of a model accounting for non-planetary effects, non-planetary($\boldsymbol{\theta}_0$), $\boldsymbol{\theta}_0$ being a real vector with n_θ components, and a sum of Keplerian or circular curves depending on five resp. three parameters, $\boldsymbol{\theta}_{K\,j} = (k_j, h_j, P_j, A_j, B_j)$ and $\boldsymbol{\theta}_{C\,j} = (P_j, A_j, B_j)$

$$\text{Keplerian}(\boldsymbol{\theta}_K) = A\dot{U}(k, h, P) + B\dot{V}(k, h, P) \tag{22}$$

$$\text{Circular}\,(\boldsymbol{\theta}_C) = A\cos\left(\frac{2\pi t}{P}\right) + B\sin\left(\frac{2\pi t}{P}\right), \tag{23}$$

where $k = e\cos\varpi$, $h = e\sin\varpi$, $\varpi = \omega + \Omega$ is the sum of the argument of periastron and right ascension at ascending node, U, V are the position on the orbital plane rotated by angle ϖ. These variables are chosen to avoid poor determination of the eccentricity and time at periastron for low eccentricities.

We compare subsequently the χ^2 of residuals of a model with n_p and $n_p + 1$ planets. In practice, one selects the tallest peak of the ℓ_1-periodogram, and uses this frequency to initialize a least-squares fit of a circular or Keplerian orbit. Then the two tallest peaks are selected and so on.

To clarify the meaning of the computed FAP, let us define the recursive periodogram, depending on a frequency ω. We denote the χ^2 of the residuals by

$$\chi^2_{K,C}\left(\boldsymbol{\theta}_0^{\rm fit}, \boldsymbol{\theta}_{n_p}^{\rm fit}, \omega\right) = \left[\boldsymbol{y} - f_{K,C}\left(\boldsymbol{\theta}_0^{\rm fit}, \boldsymbol{\theta}_{n_p}^{\rm fit}, \omega^{\rm fit}\right)\right]^T \mathbf{V}^{-1} \left[\boldsymbol{y} - f_{K,C}\left(\boldsymbol{\theta}_0^{\rm fit}, \boldsymbol{\theta}_{n_p}^{\rm fit}, \omega^{\rm fit}\right)\right] \tag{24}$$

$$\chi^2_{K,C}\left(\boldsymbol{\theta}_0^{\rm fit}, \boldsymbol{\theta}_{n_p}^{\rm fit}\right) = \left[y - f_{K,C}\left(\boldsymbol{\theta}_0^{\rm fit}, \boldsymbol{\theta}_{n_p}^{\rm fit}\right)\right]^T \mathbf{V}^{-1} \left[\boldsymbol{y} - f_{K,C}\left(\boldsymbol{\theta}_0^{\rm fit}, \boldsymbol{\theta}_{n_p}^{\rm fit}\right)\right]. \tag{25}$$

$f_{K,C}\left(\boldsymbol{\theta}_0^{\rm fit}, \boldsymbol{\theta}_{n_p}^{\rm fit}, \omega^{\rm fit}\right)$ is the model fitted depending on the non-planetary effects $\boldsymbol{\theta}_0$, the (Keplerian or circular) $\boldsymbol{\theta}_{n_p} = (\boldsymbol{\theta}_{K,C\,j})_{j=1,\ldots,n_p}$ parameters of n_p planets plus a circular or Keplerian orbit initialized at frequency ω. $\mathbf{V}$ designates the covariance matrix of the noise model ($\mathbf{V}^{-1} = \mathbf{W}^2$ with the notations above). This one is often assumed to be diagonal but this is not necessary as all the properties of those periodograms come from the fact that they are likelihood ratios. The model fit can be done linearly (Baluev 2008) or non-linearly (Anglada-Escudé & Tuomi 2012). By linear, we mean that among the five or three parameters defined in equations (22) and (23), only $(A_j)_{j=1,\ldots,n_p+1}$ and $(B_j)_{j=1,\ldots,n_p+1}$ are fitted and the non-planetary effects are modelled linearly: there exists a matrix $\boldsymbol{\Phi}$ such that non-planetary$(\boldsymbol{\theta_0}) = \boldsymbol{\Phi\theta_0}$. In the second option, the orbital elements of previously selected planets, the non-planetary effects and the signal at the trial frequency are re-adjusted non-linearly for each trial frequency.

3.7.4 FAP formulae for recursive periodograms

Recursive periodogram is a term that refers to a general concept for comparing the residuals of a model with or without a signal at a given frequency. Here we specialize the formulae we use. Denoting by $P_C(\omega)$ and $P_K(\omega)$ in the circular resp. Keplerian case,

$$P_C(\omega) = N\frac{\chi^2_C(n_p, \omega) - \chi^2_C(n_p)}{\chi^2_C(n_p)} \tag{26}$$

$$P_K(\omega) = \frac{1}{2}\left(\chi^2_K(n_p) - \chi^2_K(n_p, \omega)\right), \tag{27}$$

where $N = m - 2n_p - n_\theta$. The circular case is expression 'z_1' in equation 2 of Baluev (2008), and the Keplerian one is expression 'z' in equation 4 of Baluev (2015a). In what follows, only the circular case will be used.

The quantity we are interested in is the probability that a selected peak is not a planet. We here use the FAP as a proxy for that quantity:

$$\mathrm{FAP}(Z) = \Pr\left\{\max_{\omega\in[0,\omega_{\max}]} P(\omega) > Z \mid \text{non-planetary effects}, n_p\right\}, \tag{28}$$

where $\omega_{\max}$ is the maximum frequency of the periodogram that has been scanned. This FAP is the probability to obtain a peak at least as high as Z while there are only non-planetary effects and n_p planets. Baluev (2008) has computed tight bounds for that quantity in the case of a circular model and a linear fit (corresponding to subscript C), which we reproduce here:

$$\mathrm{FAP}(z, \omega_{\max}) \approx W\gamma\left(\frac{2z}{N_{\mathcal{H}}}\right)^{\frac{1}{2}}\left(1 - \frac{2z}{N_{\mathcal{H}}}\right)^{\frac{N_{\mathcal{H}}+1}{2}}, \tag{29}$$

where $N_{\mathcal{H}}$ is the number of degrees of freedom of the model without the sine at frequency ω, $\gamma = \Gamma(N_{\mathcal{H}}/2)/\Gamma((N_{\mathcal{H}}+3/)2)$, Γ being the Euler Γ function, and $W = \omega_{\max}\sqrt{(\bar{\boldsymbol{t}}^2 - \bar{\boldsymbol{t}}^2)/\pi}$, $\boldsymbol{t}$ being the array of measurement times and $\bar{\boldsymbol{t}}$ is the mean value of $\boldsymbol{t}$. We have also tried the exact expression of the so-called Davies bound provided by equations 8, B5 and B7 of Baluev (2008), but the results were very similar to the simpler formula. In the case of Keplerian periodogram, we used equation 21 and 24 from Baluev (2015a).

Again, we emphasize that the interest of the present method is to select candidates for future observations or unveiling signals unseen on periodograms. The FAP formulae used here do not guarantee the planetary origin of a signal. For robust results, statistical hypothesis testing (e.g. Díaz et al. 2016) can be used.

4 RESULTS

4.1 Algorithm tuning

For all the systems analysed in the following sections, the figures called ℓ_1-periodogram represent $x^\sharp(\omega)$ as defined in equation (19) plotted versus periods. The name ℓ_1-periodogram was chosen to avoid the confusion with the generalized Lomb–Scargle (GLS) periodogram defined by Zechmeister & Kürster (2009). In each case, the algorithm is tuned in the following way.

(i) The problem $(11, \mathrm{BP}_{\epsilon,W})$ is solved with SPGL1 (van den Berg & Friedlander 2008).
(ii) The solution of SPGL1 is averaged on an interval $\eta = 2\pi/(3T_{\rm obs})$ according to Section 3.5.
(iii) The grid spacing is chosen according to equation (15).

The importance of the grid span and the tolerance ϵ will be discussed in the examples.

The FAPs are computed according to the procedure described in Section 3.7.4 and are represented in Fig. 6 with decreasing FAP. The ticks in abscissa correspond to the period of the signals and the flag to their semi-amplitude after a non-linear least-squares fit.

In the following, we will present our results for HD 69830, HD 10180, 55 Cnc, GJ 876 and a simulated very active star from the RV Challenge (Dumusque et al. 2016). For each system, the GLS periodogram is plotted along with the ℓ_1-periodogram.

4.2 HD 69830

In Lovis et al. (2006), three Neptune-mass planets are reported around HD 69830 based on 74 measurements of HARPS spanning over 800 d. The precision of the measurements given in the raw data set (from now on called nominal precision) is between 0.8 and

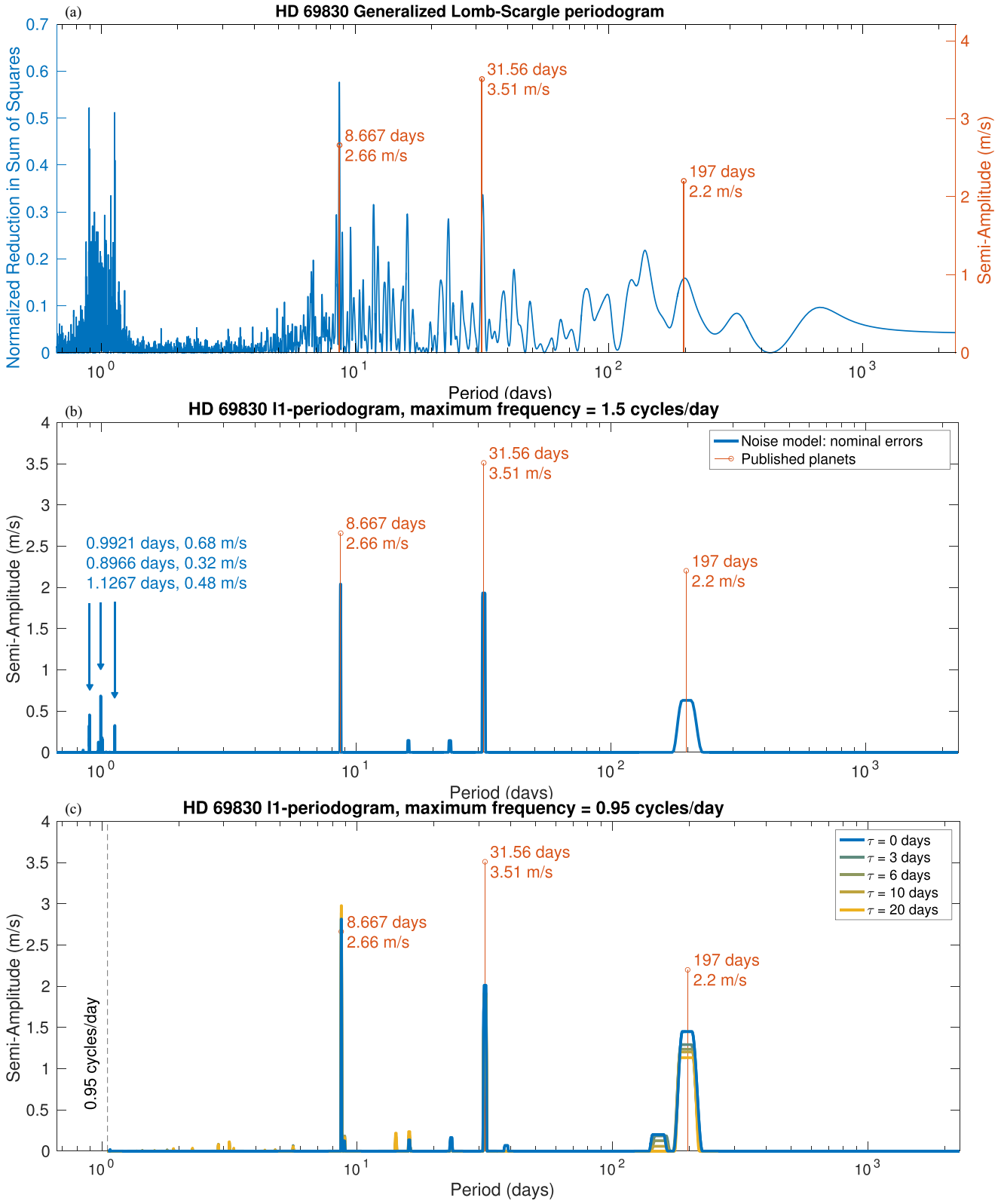

Figure 1. Generalized Lomb–Scargle periodogram and ℓ_1-periodogram of HD 69830 in blue, and published planets are represented by the red stems. The frequency spans used for panels (b) and (c) are respectively 1.5 and 0.95 cycles d^{-1}. The other signals mentioned in Section 1 are spotted by the blue arrows. For all the noise model considered for matrix $\mathbf{W}$, $\sigma_W = 0$, $\sigma_R = 1$ m s^{-1}.

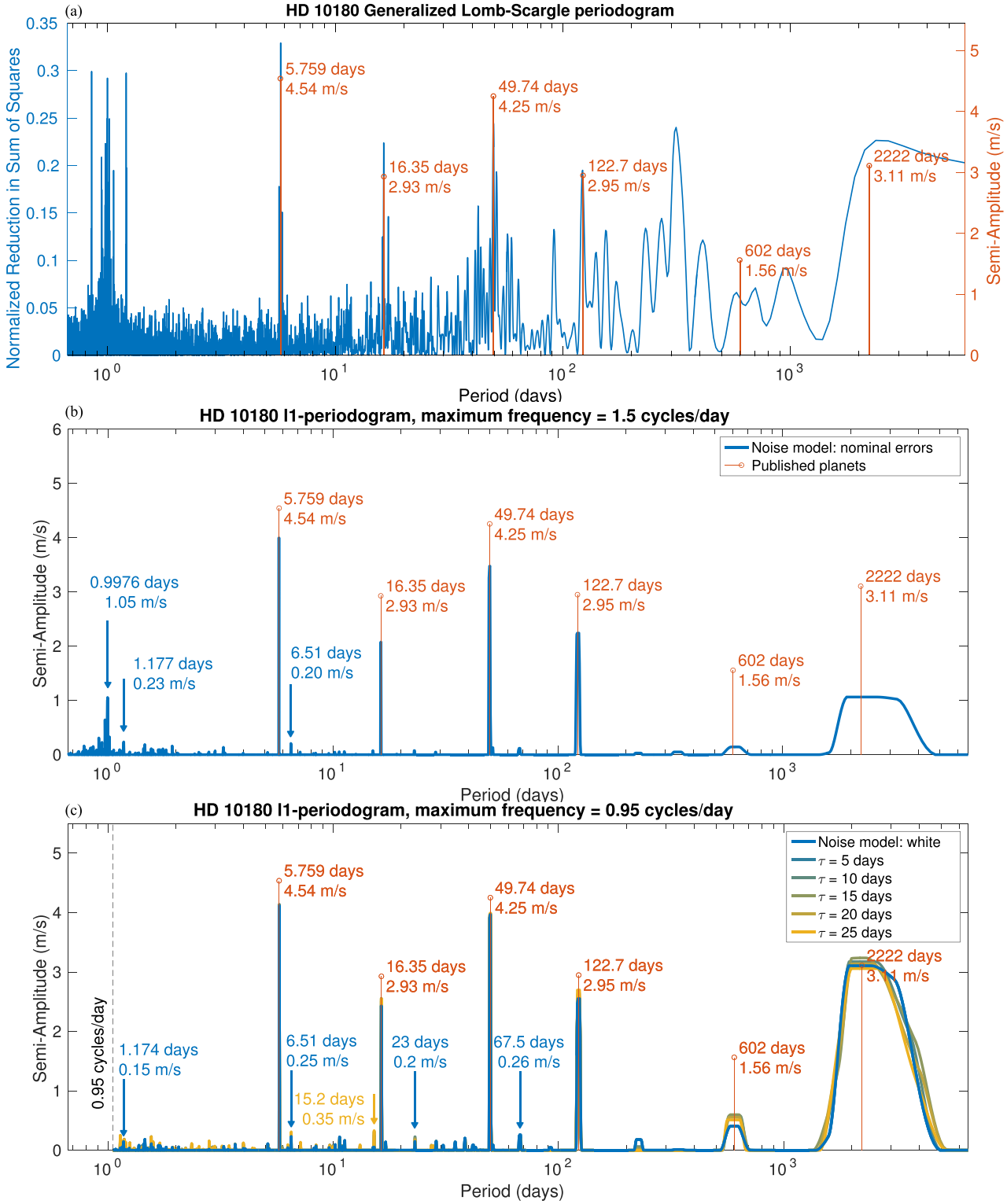

Figure 2. GLS and ℓ_1-periodograms of HD 10180 data set with mean subtracted. The red stems have the periods and amplitude of published planets. The other signals mentioned in Section 2 are spotted by the blue arrows. For all the noise model considered for matrix $\mathbf{W}$, $\sigma_W = 0$, $\sigma_R = 1$ m s^{-1}.

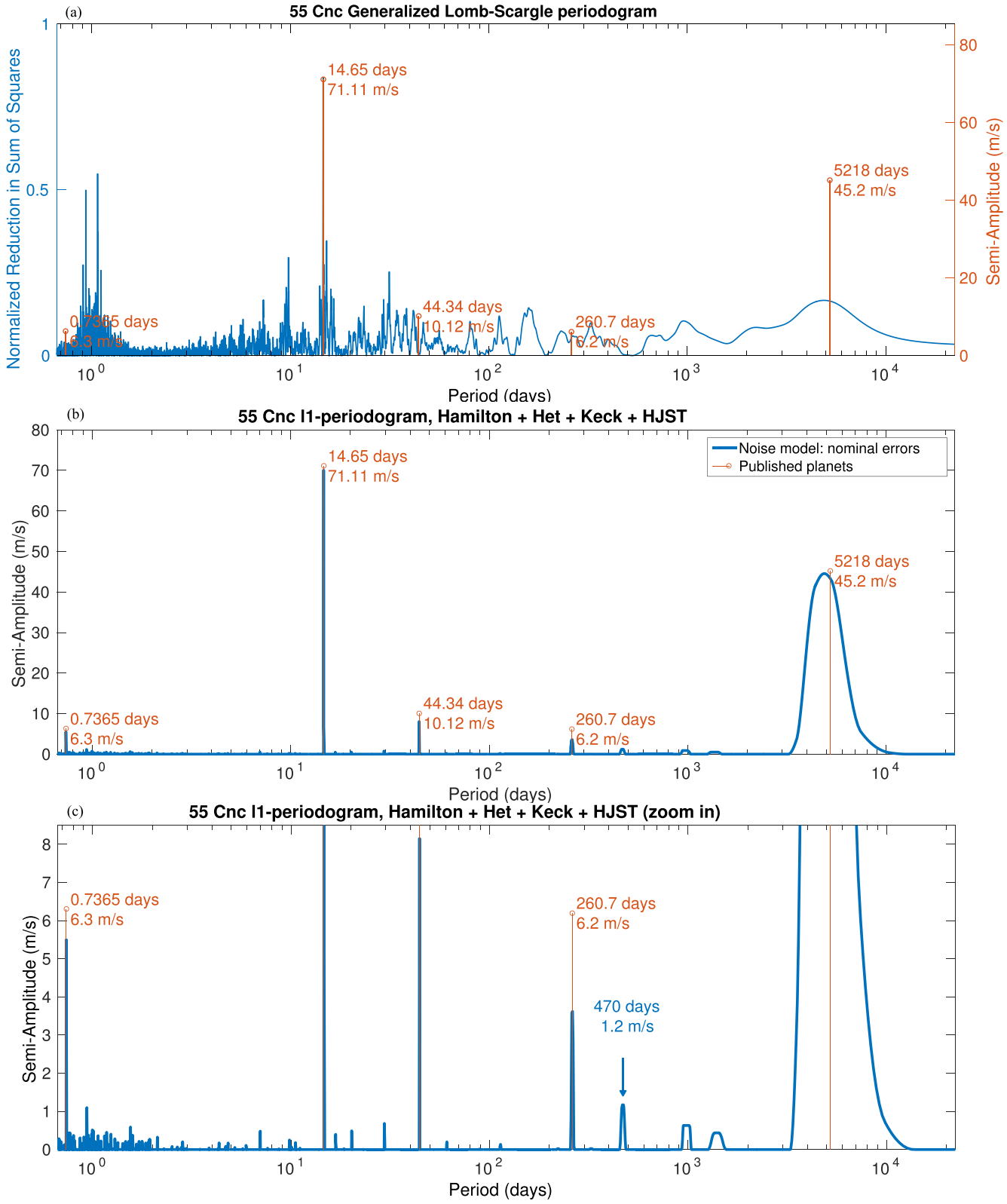

Figure 3. GLS and ℓ_1-periodograms of 55 Cnc data set with mean subtracted. The red stems have the periods and amplitude of published planets. The other signals mentioned in Section 4.4 are indicated by the blue arrows.

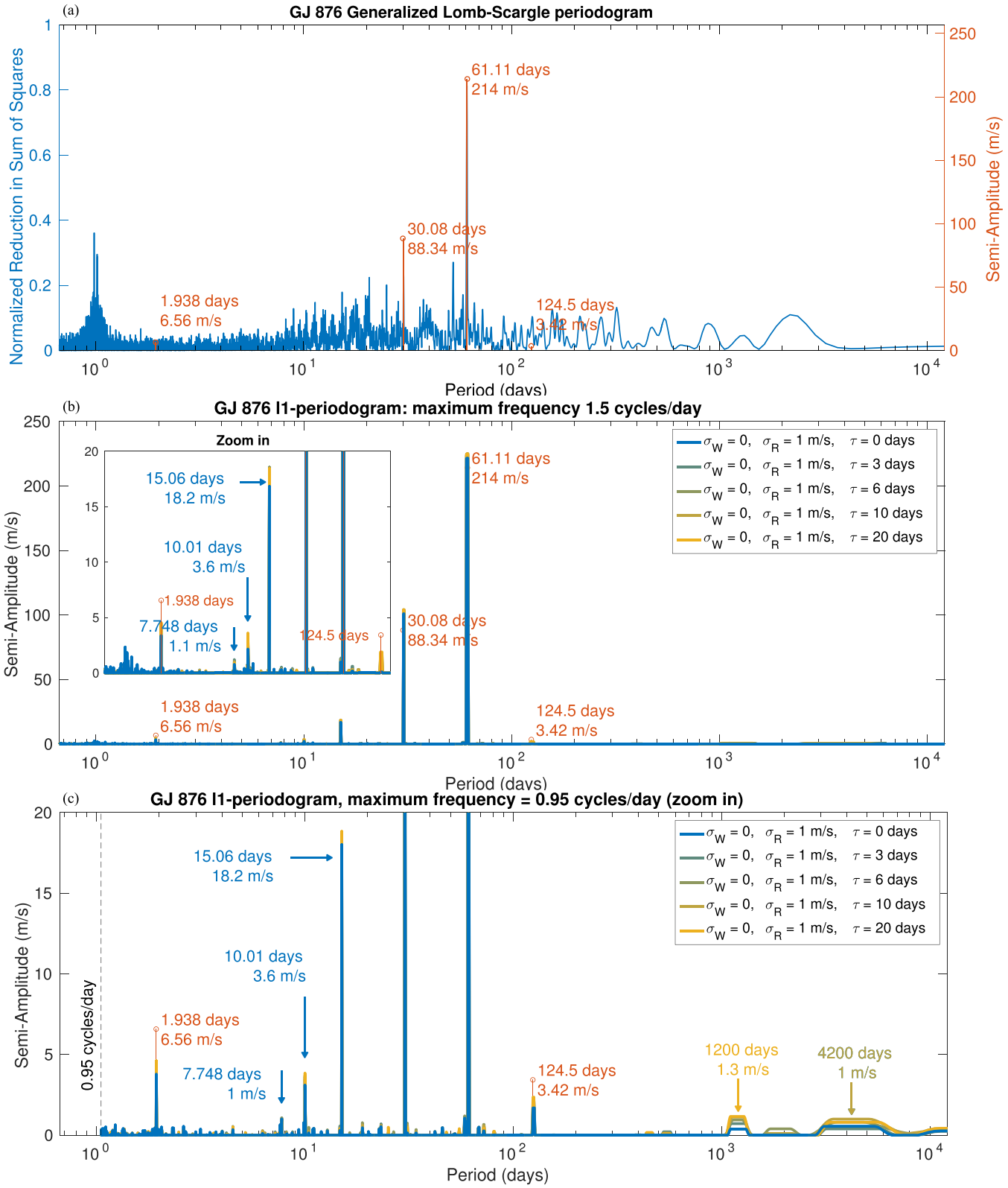

Figure 4. GLS and ℓ_1-periodograms of GJ 876 data set with means of Keck and HARPS measurement respectively subtracted. The red stems have the periods and amplitude of published planets. The other signals mentioned in Section 4.5 are indicated by the blue arrows.

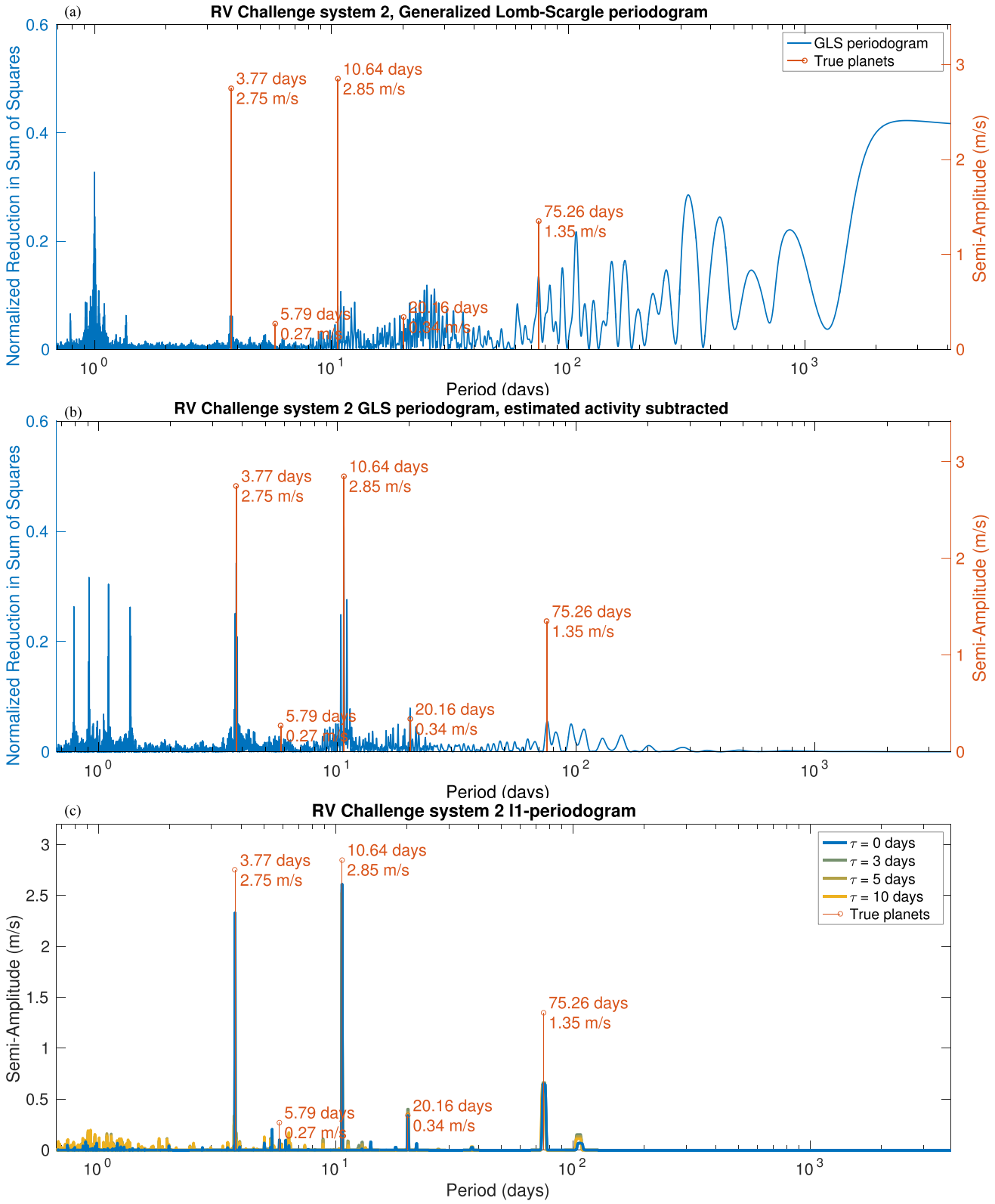

Figure 5. Top: GLS of the RV Challenge system 1 (simulated signal). Top: GLS of raw data. Middle: GLS after fitting ancillary measurements. Bottom: ℓ_1-periodogram after fitting ancillary measurements. True planets are represented by red lines.

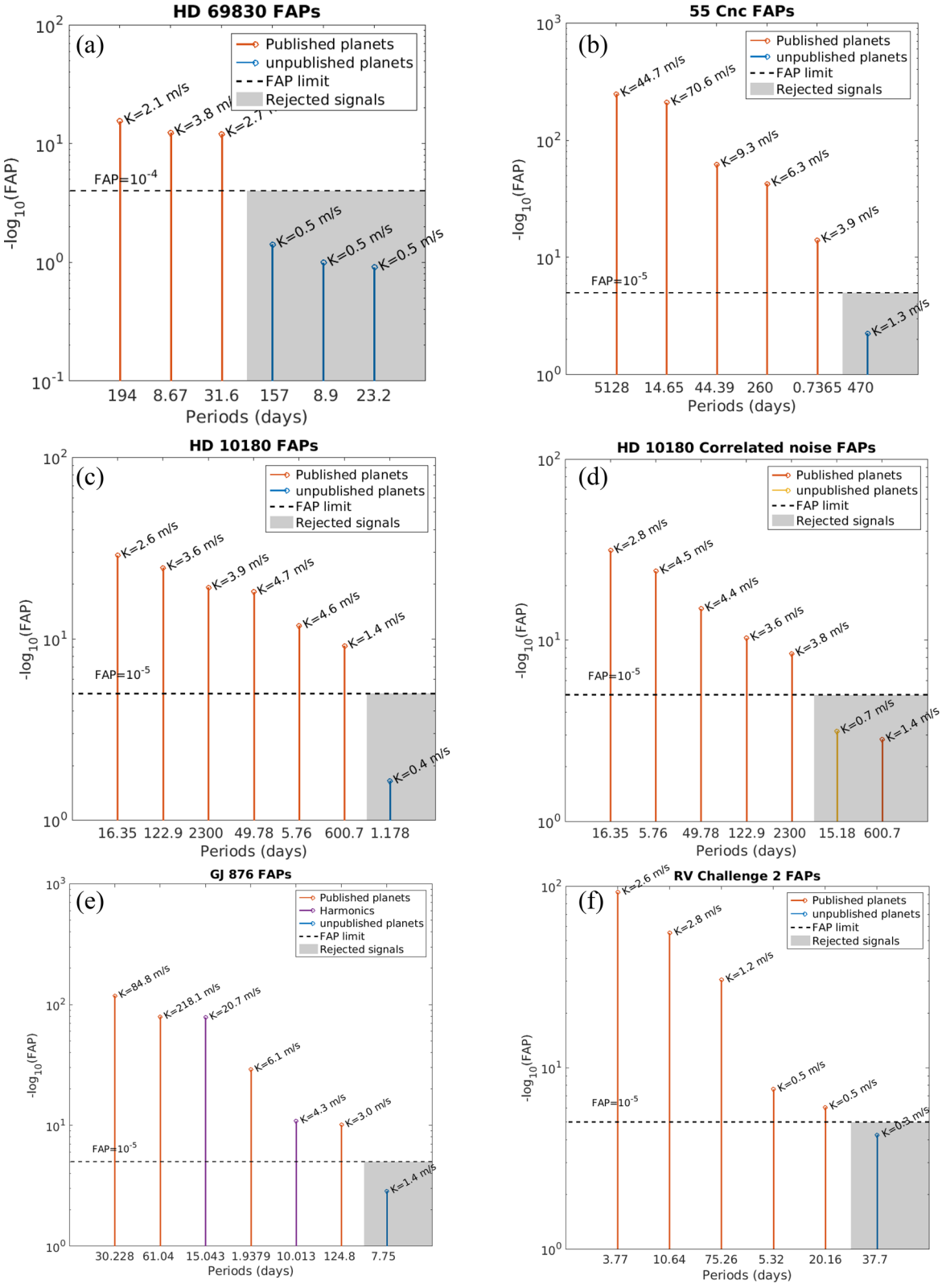

Figure 6. Peak amplitudes and associated FAPs for the four systems analysed.

1.6 m s^{-1}. The host star is a quiet K dwarf with a $\log R'_{HK} = -4.97$ and an estimated projected rotational velocity of $1.1^{+0.5}_{-1.1}$ m s^{-1}; therefore, the star jitter should not amount to more than 1 m s^{-1} (Lovis et al. 2006).

Our method consists in solving the minimization problem (11,BP$_{\epsilon,W}$) and averaging the solution as explained in Section 3.6. The resulting array $x^\sharp(\omega)$ (see equation 19) is plotted versus frequency, here giving Figs 1(b) and (c). The tallest peaks are then fed to a Levenberg–Marquardt algorithm, and the FAPs of models with an increasing number of planets are computed. We represent the FAPs of the signals when fitted from the tallest peaks to the lowest – disregarding aliases – Fig. 6(a). The FAP corresponding to an FAP of 10^{-4} is represented by a dotted line.

The values of most of the algorithm parameters defined in Section 3.5 are fixed in the previous section. In this example, the method is performed for two grid spans: 0–1.5 and 0–0.95 cycles d$_{-1}$ (Figs 1b resp. c).

We first apply the method on a grid spanning between 0 and 1.5 cycles d^{-1}. The weight matrix is diagonal, $W_{kk} = 1/\sigma_k$ (not $1/\sigma_k^2$), where σ_k is the error on measurement k. In Fig. 1(b), the peaks of published planets appear, as opposed to the GLS periodogram (Fig. 1a). However, there are still peaks around 1 d. The three main peaks in that region have periods of 0.9921, 0.8966 and 1.1267. The maximum of the spectral window occurs at $\omega_M = 6.300\,84$ rad d^{-1}. Calculating $2\pi/(\omega - \omega_M)$ yields 194.06, 8.8877 and −8.6759, respectively for $\omega = 2\pi/0.9921$, $2\pi/0.8966$ and $2\pi/1.1267$, suggesting that the short-period peaks are aliases of the true periods.

We now apply the method described in Section 3.7.4 to test the significance of the signal, obtaining Fig. 6(a). Taking 8.667, 31.56 and 197 d gives a reduced χ^2 of the Keplerian fit with three planets plus a constant (16 parameters) of 1.19, yet the stellar jitter is not included. As a consequence, finding other significant signals is unlikely.

Looking only at Fig. 1(b), whether the signal at 197 d or its alias at 0.9921 d is in the signal is unsure. We perform two fits with the two first planets plus one of the candidates. The reduced χ^2 with 0.9921 d is 1.2548, suggesting that the planet at 197 is indeed the best candidate.

Now that there are arguments in favour of a white noise and three planets, let us examine what happens when using a red noise model. The frequency span is restricted to 0–0.95 cycles d^{-1} to avoid spurious peaks (Fig. 1c). As said above, the star is expected to have a jitter in the m s^{-1} range, so we take for the additional jitter $\sigma_W = 0, \sigma_R = 1$ m s^{-1} and try several characteristic correlation time lengths $\tau = 0$, 3, 6, 10 or 20 d with definitions of equation (14). In that case, as said in Section 2.3, the estimation of the power is expected to be biased. Fig. 1(c) shows that the peaks at high and low frequencies are respectively overestimated and underestimated. We suggest the following explanation: the weighting matrix accounts for red noise that has more power at low frequencies. Therefore, the minimization of equation (5) has a tendency to 'explain' the low frequencies by noise and put their corresponding energy in the residuals.

When the signal is more complicated, there might be complex effects due to the sampling resulting in a less simple bias. This issue is not discussed in this work, but we stress that when using different matrices W, the tolerance ϵ must be tightened to avoid being too affected by the bias on the peak amplitudes.

To illustrate the advantages of our method, in Appendix C, we generate signals with the same amplitude as the ones of the present example but with periods and phases randomly selected. We show that the maximum of the GLS periodogram does not correspond to a planet in ≈7 per cent of the cases, while the maximum peak of the ℓ_1-periodogram is spurious in less than 0.5 per cent of the cases.

4.3 HD 10180

Lovis et al. (2011) suggested that the system could contain up to seven planets based on 190 HARPS measurements, whose nominal error bars are between 0.4 and 1.3 m s^{-1}. The star has an activity index $\log R'_{HK} = -5$ which lets suppose an inactive star with low jitter. In Lovis et al. (2011), the presence of the planets at 5.79, 16.35, 49.74, 122.7, 600 and 2222 d is firmly stated. Let us mention that there is a concern on whether a planet at 227 d could be in the signal instead of 600 d, as they both appear on the periodogram of the residuals and $1/227 - 1/600 + 1/365 \leq 1/T_{\rm obs}$, where $T_{\rm obs}$ is the total observation time. The possibility of the presence of a seventh planet is also discussed. After the six previous signals are removed with a Keplerian fit, the tallest peaks on the periodogram of the residuals are at 6.51 and 1.178 d (Lovis et al. 2011). They are such that $1/6.51 + 1/1.178 - 1 \leq 1/T_{\rm obs}$, so one is probably the alias of the other. The dynamical stability of a planet at 1.17 d is discussed in Laskar, Boué & Correia (2012), and its ability to survive is shown. However in our analysis, the statistical significance is too low to claim that the planet is actually in the system.

We compute the ℓ_1-periodogram for a grid span of 0–1.5 and 0–0.95 cycles d^{-1}, giving respectively Figs 2(b) and (c) (blue curve). In Appendix B, we show that when W correctly accounts for the red noise, signals might become apparent. Therefore, on the latter we also test different weight matrices. As explained in Appendix B and previous section, in that case we have to decrease $\epsilon_{\rm noise}$ and here $F_{\chi^2_m}(\epsilon^2_{\rm noise}) = 0.1$ was taken, where $F_{\chi^2_m}$ is the CDF of the χ^2 distribution with m degrees of freedom, m being the number of measurements, in accordance with the notations of Section 3.5. We note that there is a signal appearing at 15.2 d and that there is a small peak at 23 d, which is close to the stellar rotation period estimate of 24 d (Lovis et al. 2011). Whether this is due to random or not is not discussed here.

Alike the case of HD 69830, the aliases are overestimated when the frequency span is 3 cycles d^{-1}. In that case, the highest one at 0.9976 d corresponds to an alias of the 2222 d period. We will see that in the two next systems the aliases are not as disturbing, which is discussed in Section 5.2.

We now need to evaluate the significance of the peaks. The FAP test is performed for the seven highest signals, which are the published planets plus 0.177 d or 15.2 d. The latter appears for a non-diagonal weight matrix $\mathbf{W}$; therefore, when performing a Keplerian fit, the χ^2 we take is $(y(t) - \hat{y}(t))^T W^2 (y(t) - \hat{y}(t))$ with the same W, that is $\sigma_W = 0$, $\sigma_R = 1$ m s^{-1} and τ =25 d [with notations of equation (14)]. This analysis gives Figs 6(c) and (d). In both cases, the signals are below the significance threshold. It is also not clear which seventh signal to choose (Fig. 2c), but doing the analysis with other candidates as 6.51, 23 or 67.5 d does not spot significant signals either. Let us note that when choosing a non-diagonal W, the FAPs of the 16.4 and 600 d planets respectively increase and decrease. We suggest the following explanation: the noise model is compatible with noises that have a greater amplitude at low frequencies. As a consequence, the minimization has a tendency to interpret low frequencies as noise and 'trust' higher frequencies. Deciding if a signal is due to a low-frequency noise or a true planet could be done by fitting the noise and the signal at the same time.

4.4 55 Cancri

4.4.1 Data set analysis

Also known as ρ Cancri, Gl 324, BD +28°1660 or HD 75732, 55 Cancri is a binary system. To date, five planets orbiting 55 Cancri A (or HR 552) have been discovered. The first one, a 0.8 $M_{\rm j}$ minimum mass planet at 14.7 d was reported by Butler et al. (1997). Based on the Hamilton spectrograph measurements, Marcy et al. (2002) found a planet with a period of approximately 5800 d and a possible Jupiter mass companion at 44.3 d. With the same observations and additional ones from the Hobby–Eberly Telescope (HET) and ELODIE, McArthur et al. (2004) suggested that a Neptune-mass planet could be responsible for a 2.8 d period. Wisdom (2005) re-analysed the same data set and found evidence for a Neptune-sized planet at 261 d and suggested that the 2.8 period is spurious. This was confirmed by Dawson & Fabrycky (2010), which showed that the 2.8 d periodicity is an alias and the signal indeed comes from a super-Earth orbiting at 0.7365 d. The transit of this planet was then observed by Winn et al. (2011) and Demory et al. (2011), confirming the claim of Dawson & Fabrycky (2010). In the meantime, using previous measurements and 115 additional ones, Fischer et al. (2008) confirmed the presence of a planet at 261 d of minimum mass $M \sin i = 45.7$ $M_\oplus$. They also point out that in 2004 they observed two weak signals at 260 and 470 d on the periodogram. The constraints on the orbital parameters were improved by Endl et al. (2012) based on 663 measurements: 250 from the Hamilton spectrograph at Lick Observatory, 70 from Keck, 212 from HJST and 131 of the high-resolution spectrograph (HET), giving planets at 0.736 546 $\pm 3 \times 10^{-6}$, 14.651$\pm 10^{-4}$, 44.38 $\pm 7 \times 10^{-3}$, 261.2 $\pm$0.4 and 4909 $\pm$30 d. This is the set of measurements we will work on in this section. Let us mention also that Baluev (2015b) and Nelson et al. (2014) studied respectively 55 Cnc dynamics and noise correlations including additional measurements (Fischer et al. 2008).

Let us consider the set of 663 measurements from four instruments used in Endl et al. (2012). The mean of each of the four data set is subtracted and the method described in Section 2 is applied straightforwardly. Here we only display the figure obtained for a white noise model as it is essentially unchanged when correlated noise is taken into account. Fig. 3(b) shows the ℓ_1-periodogram, and Fig. 3(c) is the same figure with a smaller y-axis range. The published signals appear without ambiguity. This is somewhat surprising, as the data come from four different instruments and their respective mean was subtracted. Such a treatment is rather crude, so it shows that at least in that case the method is not too sensitive to the differences of instrumental offsets. When those are fitted with the planets found and corrected, a 365 d periodicity clearly appears on the ℓ_1-periodogram.

The FAPs computed following the method outlined in Section 3.7.4 are significant (see Fig. 6b). The sixth highest peak is at 470 d, the FAP of which is too low to claim a detection. Interestingly enough, a signal at this period was mentioned by Fischer et al. (2008). We will see in the next section that this one is already seen in 2004, and probably due to the different behaviour of the instruments at Lick and HET. The presence of a signal at 2.8 and 260 d in early measurements is also discussed.

4.4.2 Measurements before 2004: no planet at 2.8 d nor 470 d but visible 55 Cnc e and f

The 55 Cnc system has several features that are interesting to test our method. There has been some false detections at 2.8 d, and among candidate signals, one was confirmed (260 d) and one was not (470 d). We now have at least 663 reliable measurements that are very strongly in favour of five planets. As a consequence, the method can be applied on a shortened real data set with specific questions in mind, while being confident about what really is in the system. We will see that the use of the ℓ_1-periodogram could have helped detecting the true planets based on the 313 measurements considered in McArthur et al. (2004). These ones are from Hamilton spectrograph at the Lick Observatory, the HET and ELODIE (Observatoire de Haute Provence). We also show that the signal at 0.7365 d (55 Cnc e) was detectable on the separate data sets from Lick or from HET available in 2004.

Our method is first applied to the three data sets at once, the means of which were subtracted, which gives the lighter blue curve in Fig. 7(a). The true periods appear, although the 260 period is very small and there are peaks at 470, 1314 and 2000 d (the other features of the figure will be explained later). We then consider the three data sets separately; the figure obtained is displayed in Fig. 7(b). The fact that the ℓ_1-periodograms of each of the three instruments span on different length is due to the fact that they do not have the same observational span. As the moving average on the result of SPGL1 is $2\pi/3T_{\rm obs}$, it is wider when the total observation time $T_{\rm obs}$ is small. The 14.65 and long periods are seen for each data set, but the 0.7365 and 44.34 d periodicities are not seen for the ELODIE data set. Interestingly, HET ℓ_1-periodogram displays a periodicity close to 260 d. However, one cannot claim a detection at this period in HET data, as those only span on 180 d, any period longer than the observation timespan is very poorly constrained. Furthermore, the period at 2.8 d is not seen in any data set. The closest one would be a peak at 2.62 d obtained with ELODIE data, which was checked not to be significant. The 470 d periodicity does not appear either. We show in the next paragraph that this is likely due to the velocity offset between Lick-Hamilton and HET data sets. Let us point that CLEAN (Roberts et al. 1987) or frequency analysis (Laskar 1988; Laskar et al. 1992, see Fig. 8) also allows us to retrieve the 0.7365 periodicity, which basically means that the strongest peak of the residual was already this one in 2004.

To compute the significance, the method of Section 3.7.4 is applied to the Lick and HET data separately. The FAPs are computed for circular models with an increasing number of planets whose periods correspond to the subsequent tallest peaks of the ℓ_1-periodogram. Here, as the data come from different instruments we add to the model three vectors $1_{\rm Lick}(t)$, $1_{\rm Elodie}(t)$ and $1_{\rm HET}(t)$, where $1_I(t) = 1$ if the measurement at time t is made by instrument I, $1_I(t) = 0$ otherwise. In the case of Lick data, there is a peak of 6 m s^{-1} at 1.0701 d, but this one can be discarded as it is an alias of the 14.65 d periodicity. In both HET and Hamilton data, the 0.7365 periodicity is significant (Figs 7c and d). Also, one sees a significant long period in both cases (respectively 8617 and 5212 d). The HET data set spans on 170 d, so in this case one can only guess that there is a long-period signal. Finally, when combining the two data sets, the 470, 2150 and 1314 d periodicities become insignificant.

The difference in zero-points of the three instruments has a signature on the ℓ_1-periodogram. Indeed, in problem (11,BP$_{\epsilon,W}$), the signal is represented as a sum of sinusoids. The algorithm could then attempt to 'explain' the bumps in velocity that occur when passing from one instrument to the other by sines. The previous analysis ensures the presence of four periodicities in the signal: at $\approx$14.65 d, 44.34, 5000 and 0.7365 d. The fit with these four periods plus the vectors $\boldsymbol{I_I}(t)$ gives coefficients of the latter $\alpha_{\rm Lick}$, $\alpha_{\rm Elodie}$ and $\alpha_{\rm HET}$. The vector $\alpha_{\rm Lick}\boldsymbol{I}_{\rm Lick}(t) + \alpha_{\rm Elodie}\boldsymbol{I}_{\rm Elodie}(t) + \alpha_{\rm HET}\boldsymbol{I}_{\rm HET}(t)$ is subtracted from the raw data. The ℓ_1-periodogram of the residuals

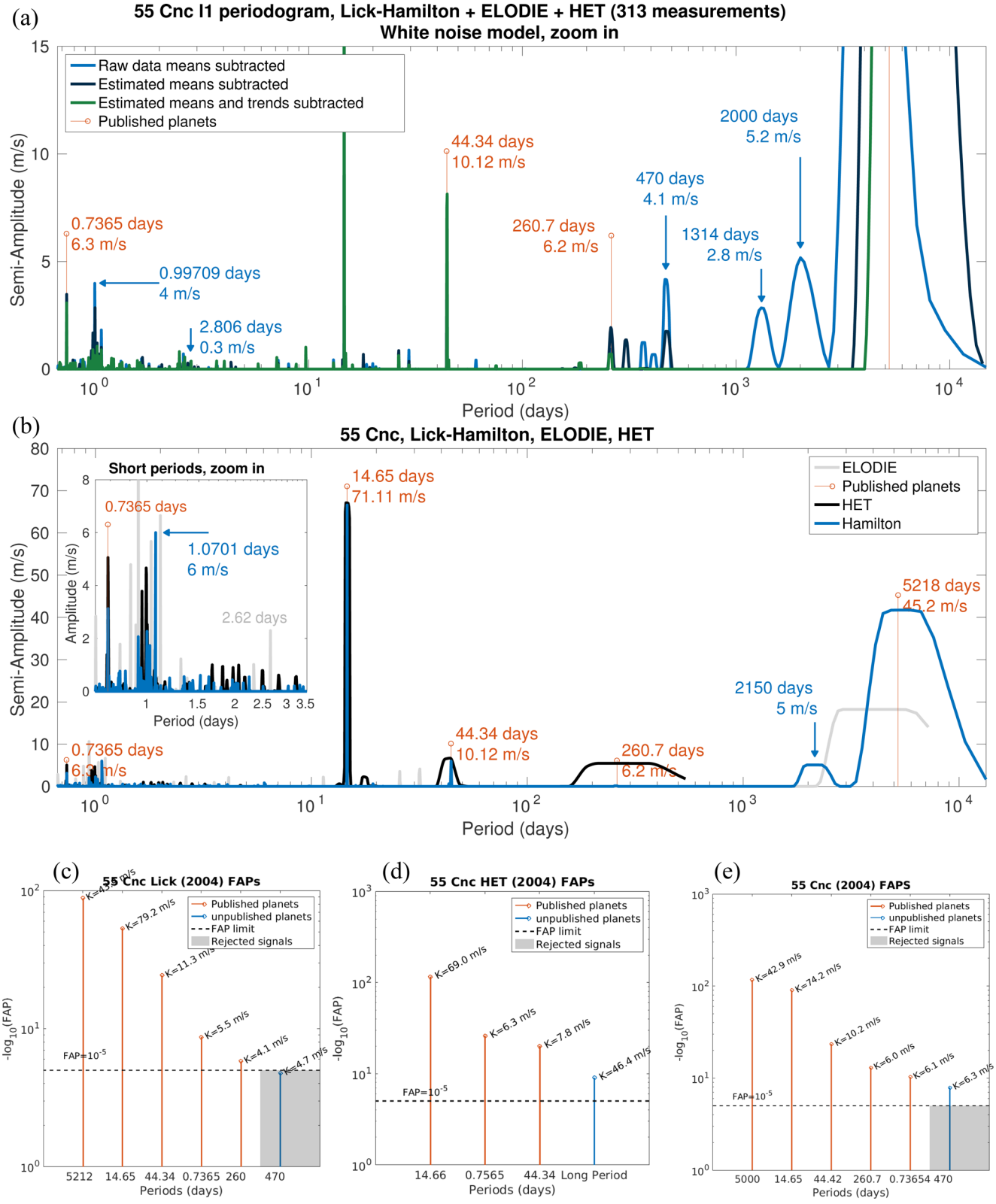

Figure 7. ℓ_1-periodogram of 55 Cnc, using measurements from the Lick-Hamilton, ELODIE spectrograph (Observatoire de Haute Provence) and HET telescope.

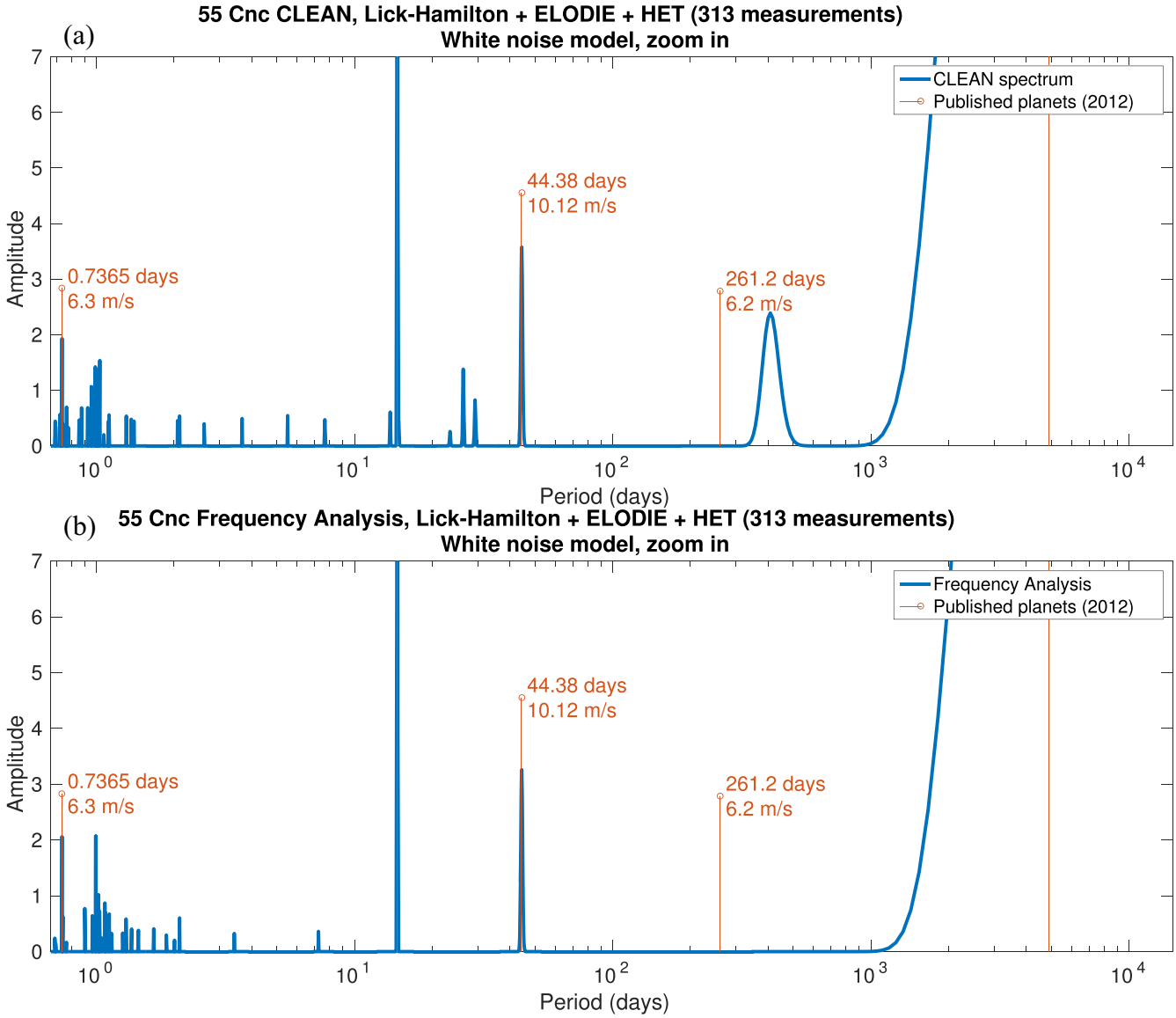

Figure 8. (a) CLEAN spectrum of 55 Cnc with the data available in 2004. (b) Frequency analysis of the same data.

is computed, which gives the dark blue curve shown in Fig. 7(a). The 2000 and 1314 periods disappear and the 470 d peak decreases. Interestingly enough, the fifth tallest peak (except the 0.99 709 d alias) becomes 260 d, which was suggested by Wisdom (2005) and confirmed by Fischer et al. (2008) and Endl et al. (2012), but it does not appear on the CLEAN spectrum nor the frequency analysis (Figs 8a and b).

We now fit the model with five planets along with the $\boldsymbol{I}_I$ vectors and trends for each instrument, which are vectors $\boldsymbol{t}_I$ such that $\boldsymbol{t}_I(t) = t$ and 0 elsewhere if the measurement at time t is done by the instrument I. The vector $\sum \alpha_k 1_{I_k} + \beta_k \boldsymbol{t}_{I_k}$ is subtracted from the raw data, and we compute again the ℓ_1-periodogram (Fig. 7a, green curve). This time, the 470 d periodicity disappears, suggesting – though not proving – that it is due to a difference in behaviour between the instruments. The fact that the 470 d signal disappears just shows its presence depends on the models of the instruments. The same analysis on Lick and HET data altogether shows the same features at 470 d; therefore, we exclude the possibility that it is due to the lesser precision of ELODIE.

The analysis by Wisdom (2005) does not use ℓ_1 minimization to unveil the 260 d periodicity (55 Cnc f). We tried to reproduce a similar analysis 'by hand' on the same data set, namely the one of McArthur et al. (2004). The rationale is to determine if it was easy to make 55 Cnc f appear with an analysis more conventional than the ℓ_1-periodogram. Also, the short-period planet can be injected at 0.7365 d, not ≈2.8 d as it was then. We found that the size of the peak in the residuals at 260 d depends on the initialization of the fits, both with classical and recursive periodograms. While in most cases the 260 periodicity does appear in the residuals, it sometimes coexists with peaks of similar amplitude. Interestingly enough, an analysis of Lick-Hamilton and HET data sets by recursive periodograms suggests that the periods estimated by HET are shifted to longer ones with respect to Lick ones. We found that adding the periods 14.8, 15 000 ($1/14.65 - 1/14.8 \approx 1/5000 - 1/15\,000$) to those of the four planets and a 2500 one (probably due to a harmonic of the 5000 d periodicity) makes the 400 (seen on the CLEAN spectrum in Fig. 8a) and 470 periodicity disappear, and the 260 d peak appears very clearly. As the data come from

an older generation of spectrographs, one could expect complicated systematic errors. Again, this discussion focuses on the possibility of seeing the 55 Cnc f in 2004; we do not raise the question of its existence, well established by the subsequent measurements.

Finally, we perform the FAP test on the data from the three instruments (see Fig. 7e). The model is made of Keplerians plus the 1_I vectors. The four significant signals in each data set are still significant. The 260 d periodicity is significant as well. This analysis shows that both the 0.7365 and 260 d periodicities were already present in the data. Long periods might be due to instrumental effects; therefore, the planetary origin of the 260 period could have been subject to discussion. In contrast, it seems hard to explain a steady 0.7365 d periodicity with a non-planetary effect. Let us also note that our χ^2 minimization algorithm and the one used by Wisdom (2005) are different, the latter being more sophisticated, which might account for discrepancies between our analyses.

4.5 GJ 876

4.5.1 Previous work

The GJ 876 host star is one of the first discovered multiplanetary systems. First, two giant planets at 30 and 61 d were reported by Marcy et al. (1998) and Delfosse et al. (1998). Subsequently, Rivera et al. (2005) find a short-period Neptune at 1.94 d and a Uranus-mass planet at 124 d (Rivera et al. 2010).

The giant planets are close to each other and in 2:1 resonance, therefore we might expect visible dynamical effects. Indeed, Correia et al. (2010), Baluev (2011) and Nelson et al. (2016) perform four-body Newtonian fits which give a χ^2 of the residuals smaller than a Keplerian fit. The dynamical fits also allow one to have constraints on the inclinations, therefore on the true masses of the planets. Furthermore, Baluev (2011) shows that the maximum of a posterior likelihood including a noise model as the one used here (equation 14) occurs at $\sigma_W = 1.31$ m s^{-1}, $\sigma_R = 1.8$ m s^{-1} and $\tau = 3$ d.

Jenkins et al. (2014) take a different approach and search for sine functions in the signal. They claim that six significant sinusoidal signals are in the data. The following discussion first confirms these results. Secondly, we investigate the origins of the additional two signals and find that they are likely to be due to the interactions between the giant planets.

4.5.2 Six significant sines

Jenkins et al. (2014) analyse the GJ 876 data by aiming at solving the problem (1), which they call minimum mean squared error. To do so, the phase space is explored with an iterative arborescent method. They find the following periods: 61.03±3.81, 30.23±0.19, 15.04±0.04, 1.94±0.001, 10.01±0.02 and 124.69±90.04 d. To compare our results with Jenkins et al. (2014), the significance of the signals is tested with FAPs as previously. We use different weight matrix models according to equation (14) and two grid spans: 1.5 and 0.95 cycles d^{-1} (see Figs 4b and c). In Fig. 4(c), we see that the six tallest signals correspond to the periods we expect. Depending on the noise model, the seventh tallest peak varies. We compute the FAP test for 7.748, 1200 or 4200 d as candidate seventh planets, respectively, with the $\mathbf{W}$ matrix yielding their greatest amplitude. In Fig. 6(e), we display the result for 7.748 d but in other cases the signals are not significant. Let us still point out that in the case of $\tau = 6$ d, initializing a 4200 d periodicity, after the non-linear fit we obtain a 4862 d periodicity which has an FAP of 0.0007. This one is close to the total observation timespan (4600 d). Therefore, it is hard to determine what could be its cause.

Before discussing the origin of these signals, we wish to comment on the behaviour of the ℓ_1-periodogram towards the 124 d periodicity. Indeed, in the case of the 1.5 cycles d^{-1}, this one has the same order of magnitude as the tallest alias in the 1 d region (at 0.9812 d, alias of the 61 d periodicity). Furthermore, the peak becomes visible only for non-diagonal weight matrix $\mathbf{W}$, while a white noise model is sufficient to see it when using a shorter grid (Fig. 4c). To understand this feature, we argue as follows. There are three effects against finding the correct planets: the red noise (Baluev 2011), the uncertainties on the two instrumental means and the inner faults of our method. The persistence of aliases at 1 d indeed shows that the recovery of the true signals is more difficult when considering a grid Ω where some of the frequencies are very correlated. We also computed the ℓ_1-periodogram when the mean of each instrument is corrected after the orbital parameters fit, as done in Section 4.4.2. In that case, the 124 d periodicity does appear and the aliases are reduced. We suggest the following explanation: when at least one of the three obstacles is correctly taken into account, the method is sufficient. When the three are ignored, their joint effect is deadly to our ability to recover the correct planets.

4.5.3 Signals at 10 and 15 d

Now that the six sines are seen in the signal, we show that the peaks at 15.06 and 10.01 d are due to the dynamical interactions.

We perform the same four-body fit of GJ 876 with the same method as Correia et al. (2010). This one includes 25 parameters: the mass of the star, a velocity offset, the mass of the planets, for the smallest planets: period, semi-amplitude, eccentricity, argument of periastron and initial mean anomaly. For the giant planets at 30 and 61 d, the inclination is also a free parameter.

A planetary system with the orbital elements found by the least-squares fit is simulated on 100 years for the two giant planets and the four planets at once. The frequency analysis (Laskar 1988, 1993; Laskar et al. 1992) is then performed on the resulting time series of the star velocity along the x-axis. We find that 15.06 and 10.01 periods appear and are a combination of the fundamental frequencies. Denoting by ω_P the frequency of a planet of period P, we have $\omega_{15} = 3\omega_{30} - 2\omega_{60}$ and $\omega_{10} = 5\omega_{30} - 4\omega_{60}$, both in the two-planet and four-planet cases. We also performed another test: if we adjust the two giant planets with a dynamical fit, then the peaks at 15.06 and 10.01 d are not seen on the residuals. This agrees with the analysis of Nelson et al. (2016), where they discuss the possibility that the signals at 10.01 and 15.06 d could be due to additional planets, and find it unlikely. They compute the evidence ratio of Newtonian models with four and five planets, $\Pr\{\boldsymbol{y}|5\text{ planets}\} / \Pr\{\boldsymbol{y}|4\text{ planets}\}$, and find that it is not higher than the threshold we chose. The difference between ω_{15} and the first harmonic of the planet gives an estimate of the frequency of precession of the periastron of the inner orbit; we find $2\pi(1/\omega_{15} - 2/\omega_{30}) \approx 8.77$ yr, which is consistent with the estimate of Correia et al. (2010, $g_2 = 8.73$ yr, table 4).

To obtain the expressions of ω_{15} and ω_{10}, we used frequency analysis. This could be puzzling as the present work defines a method to retrieve the frequencies in the signal. The rationale is that we do the frequency analysis on a numerical integration; therefore, we have tens of thousands of points available. Frequency analysis has been used in that situation for years and is known to be fast and robust. We double checked the results by computing the ℓ_1-periodogram on a thousand points from the simulation (handling as many as the frequency map analysis is too long for now); the periods at 15.06 and 10.01 do appear very clearly.

4.6 Very active star (simulated signal)

The examples above concern rather quiet stars, where the noise can be modelled by Gaussian time series. However, in some cases, the stellar activity does not have a known Gaussian signature. The method described here is not yet adapted to handle such situations. In this section, we show that the problem can be circumvented, provided there are enough measurements.

We exploit the fact that stellar noise can be correlated with the bisector span (Queloz et al. 2001), the full width at half-maximum (FWHM) and the $\log R'_{HK}$. This correlation has been used for example in Meunier, Lagrange & De Bondt (2012), which shows that the detection threshold limit improves by an order of magnitude by testing the correlation between the radial velocity and ancillary measurements. They compute the correlation of the periodograms of radial velocity measurements and bisector span, but a correlation in the frequency domain is also visible in the time domain, as the Fourier transform contains the same amount of information as the original time series. Here we take an approach similar to Melo et al. (2007), Boisse et al. (2009) and Gregory (2016) in so far as we use the ancillary measurements as proxies for estimating the activity induced signal. Here, we simply fit and remove the three ancillary measurements from the data and then use the method described above on the residuals. To compute the FAP, we use a model of the form $A\mathrm{FWHM} + B\mathrm{bisector} + C \log R'_{HK} + \mathrm{Circ}(k, h, P, D, E)$, Circ denoting a circular model as defined in Section 3.7.3. The validity of this approach is discussed in Appendix D.

The data set used is taken from the RV Fitting Challenge (Dumusque 2016; Dumusque et al. 2016). In this challenge, 15 systems were simulated with a red noise component taken from observations of real stars plus activity simulated via SOAP 2 (Dumusque, Boisse & Santos 2014). Here we consider the system number two of the challenge. The data set is made of 492 measurements, and the mean precision is 0.67 cm s^{-1}. The first step of the processing is to fit a linear model made of the ancillary measurements, an offset, a linear and a quadratic trend (six parameters). Secondly, we compute the ℓ_1-periodogram for different weight matrices, which gives Fig. 5(c). The GLS periodogram is also computed before and after the fit of the six parameters for comparison (Figs 5a and b).

We find without ambiguity the three planets whose semi-amplitude is above 1 m s^{-1}, and also the 20.16 d periodicity. The planet with the smallest amplitude does not appear clearly, but there is a peak at 5.4 d which seems to be significant. In fact, the spectral window is such that 5.4 d is an alias of $5.32 = 10.64/2$ d, and corresponds to the first harmonic due to eccentricity. This feature seems to be due to an error in the noise model. When accounting for a red noise effect, the relative amplitudes of 5.32 and 5.4 change in favour of 5.32 d. This effect is also observed on the recursive periodograms which are not represented here for the sake of brevity. One can see a peak at 6.25 d which grows stronger as the characteristic correlation time of the noise model increases. This coincides with the fourth harmonic of the rotational period and is therefore not surprising.

5 DISCUSSION

5.1 Summary

The present work was first devised to overcome the distortions in the residual that arise when fitting planets one by one. It is compatible with the assumption that the noise is Gaussian and correlated through the weighting matrix $\mathbf{W}$. One of the main advantages of the method is that, as opposed to global χ^2 minimization, the minimization problem (11,$\mathrm{BP}_{\epsilon,W}$) is convex therefore quicker to solve. On our workstation (Intel Xeon CPU E5-2698 v3 at 2.30 GHz), it takes typically 30 s to 10 min to obtain (resp. for HD 69830, 74 measurements and 55 Cnc, 663 measurements). The speed here depends mainly on three parameters: the number of observations m, the number of columns of matrix $\mathbf{A}$ (see Section 3.3), n, and the precision wanted in output, tol (see Section 3.2). The SPGL1 algorithm used to solve equation (11,$\mathrm{BP}_{\epsilon,W}$) relies on a Newton algorithm; therefore, its complexity is $O(\log(p)F(p))$, where $p = 10^{-\mathrm{tol}}$ is the number of significant digits desired and $F(p)$ the cost of evaluating the objective function to p digits accuracy. The most expensive steps of the evaluation are a matrix vector product and a projection on to a convex set (see van den Berg & Friedlander 2008), which have a respective complexity of $O(mn)$ and a worst case complexity of $O(n \log n)$. The post-processing operation also is in $O(mn)$. This overall should amount asymptotically to complexity $O(mn)$, similarly to the Lomb–Scargle periodogram. Its complexity is in $O(mn)$ if there are m measurements and n frequency scanned. The constants are however different.

Furthermore, our method does not require the number of planets as input parameter and offers a graphic representation of the information content of the signal. However, the statistical properties of the solution are not as easy to interpret as in the case of a global least-squares minimization. Considering that the method presented here is in its infancy, comparing its merits to other techniques is left for future work. Here, we will only stress that the ℓ_1- and GLS periodogram are tools are of different levels, and we do not advocate to give up the latter.

We will confine ourselves to addressing some internal issues of our method. Ultimately, we would like to know if there is a way to determine which peaks are to be associated with planets. As the present paper is concerned with unveiling the periodicities in the signal but not their origins, we will address a simpler question: assuming that the signal is only made of sines plus a Gaussian noise, are there risks to see spurious peaks on the ℓ_1-periodogram?

Unfortunately, the answer is yes, as we have seen in the previous examples. The method is in particular sensitive to the aliases due to the daily repetition of the measurements: spurious peaks are especially present around 1 d periods. To shed some light on this problem, the following questions will be briefly discussed in the two next sections.

(i) Are spurious peaks to be expected from the theoretical properties of the method or from its implementation?

(ii) If they are to appear anyway on the ℓ_1-periodogram, is there a way to spot them?

5.2 Mutual coherence

To test if the algorithm behaves appropriately, we reason as follows. Considering a set of observational times $\boldsymbol{t} = t_1, \ldots, t_m$, a linear combination of p pure sine signals $y(t_k) = a_1 \cos(\omega_1 t_k + \phi_1) + \cdots + a_p \cos(\omega_p t_k + \phi_p)$ is generated with uniformly distributed phases ϕ and various amplitudes. For any tolerance ϵ, the SPGL1 algorithm must give a solution $\boldsymbol{x}^\star$ (see equation (11,$\mathrm{BP}_{\epsilon,W}$)) such that $\|\boldsymbol{x}^\star\|_{\ell_1} \leqslant |a_1| + \cdots + |a_p|$, as obviously $\boldsymbol{y}(\boldsymbol{t})$ belongs to the set of signals u verifying $\|\boldsymbol{u} - \boldsymbol{y}(\boldsymbol{t})\|_{\ell_2} \leqslant \epsilon$. To test if SPGL1 gives the best solution, we take the measurement dates of HD 69830 and generate three pure cosine functions of one amplitude whose frequencies are in the grid. They are fed to the SPGL1 solver for $\epsilon = 0.01$ and W equal to the identity matrix. The solution $x^\star$ to equation (11,$\mathrm{BP}_{\epsilon,W}$) must verify

Table 1. Maximum amplitude of the spectral window in the 1 cycle d^{-1} and 1 cycle yr^{-1} for the examples considered here.

	≈1 cycle d^{-1}	≈1 cycle yr^{-1}
HD 69830	0.926	0.600
HD 10180	0.949	0.703
55 Cnc	0.822	0.557
GJ 876	0.732 46	0.501
RV Challenge 2	0.870	0.800

$\|\boldsymbol{x}^\star\|_{\ell_1} < 3$ as the original signal is not noisy. The test is performed for a thousand set of three frequencies randomly selected on the grid. We find that the average ℓ_1 norm of the solution is 3.26, suggesting that the algorithm could be improved.

Secondly, in the discrete case (problem 5), there are theoretical guarantees on the success of the recovery if the mutual coherence of the dictionary is sufficiently small (Donoho 2006). This one is defined as the maximal correlation between two columns a_j and a_k of the dictionary A,

$$\mu = \max_{\substack{k=1,\dots,n \\ j=1,\dots,n \\ j\neq k}} |\langle \boldsymbol{a_k}, \boldsymbol{a_j} \rangle|. \tag{30}$$

In the case of a dictionary such that $\boldsymbol{a}_k = \mathrm{e}^{\mathrm{i}\omega_k \boldsymbol{t}}$, taking the convention $\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle = \boldsymbol{a}_k^* \boldsymbol{a}_j$, where the superscript $*$ denotes the conjugate transpose,

$$|\langle \boldsymbol{a_k}, \boldsymbol{a_j} \rangle| = \left| \sum_{l=1}^{m} \mathrm{e}^{-\mathrm{i}(\omega_k - \omega_j)t_l} \right| \tag{31}$$

that is the spectral window in $\omega_k - \omega_j$. As a consequence, the method cannot resolve very close frequencies due to their high correlation. More importantly, aliases are still a limitation – though not as much as in iterative algorithms in general (Donoho et al. 2006), see also Appendix C. This feature is responsible for the aliases that still appear around 1 d, where there is generally a strong alias due to the sampling constraints. The problem tends to get worse as the maxima of the spectral window increase. Aliases are higher relative to the true peaks for HD 69830, HD 10180 and the separate sets of 55 Cnc than GJ 876 (see Figs 1, 2, 3, 4, 7 and Table 1).

5.3 Spotting spurious peaks

We know that the theoretical obstacle for a good recovery is correlation between the elements of the dictionary. If a frequency ω_0 truly is in the signal, it is expected to cause significant amplitudes at $\omega_0 + \omega_k$, where the ω_k are maxima of the spectral window. So if two peaks at frequencies ω_1 and ω_2 are seen on the ℓ_1-periodogram and the spectral window has a strong local maximum close to $\omega_1 - \omega_2$, one can suspect that one of the two peaks is spurious.

5.4 When to use the method?

We consider the general problem of finding the frequencies of a signal made of several harmonics (the multitone problem). It seems natural – though not mandatory – to try to find the global minimum for a given number of sinusoids, and possibly additional parameters such as the offset or a trend. We do not know a priori the number of sinusoids in the signal. Ideally, we would like to solve the global minimization (1) for any number of sines inferior to the number of measurements and regarding their amplitudes, which seem to be truly in the signal. The approach consisting in using grids has a computational cost growing exponentially with the number of frequency. Therefore, strategies must be found to estimate a reliable solution to this problem. The recursive periodogram (Angladа-Escudé & Tuomi 2012), the treillis approach (Jenkins et al. 2014) or the super-resolution methods (Candès & Fernandez-Granda 2014; Tang et al. 2013b) can be viewed as a way to approximate equation (1) and selecting the relevant number of frequencies at the same time. These have the advantage of not being bothered by the ℓ_1 norm minimization, which biases downwards the amplitude of the signal. Even more, the bias becomes more complicated when using a correlated noise model.

The most interesting use of the ℓ_1-periodogram seems to be as a complement to the classical periodogram: it gives a much clearer idea of the number of spikes and their significance. If the peaks spotted by the ℓ_1-periodogram yield a χ^2 of the residuals consistent with the noise assumptions as in HD 69830, then it is likely that there are not many more signals. To check that there are not very high correlations between signals, one can use the spectral window. Furthermore, we have exhibited in Appendix C1 examples where the main peak of the classical periodogram is spurious while ℓ_1 minimization (5) avoids selecting the first spurious peak. Such an example was also presented in Bourguignon et al. (2007). Those findings are consistent with the claims of Donoho et al. (2006): the ℓ_1 method is more reliable in general than orthogonal matching pursuit. A failure of the ℓ_1-periodogram is also informative, as shown in Fig. C2 of Appendix C1. If there still is a forest of peaks below a certain amplitude, it might indicate that the signal is noisy, possibly that noise is higher than expected or non-Gaussian. This means that the set of observations requires a more careful analysis. To sum up, the ℓ_1-periodogram can yield an estimation of the difficulty of the system; in some cases, it is a shortcut to random searches and its use decreases the chance of being misled by a spurious tallest peak.

6 CONCLUSION

The aim of the present paper was to produce a tool for analysing radial velocity that can be used as the periodogram but without having to estimate the frequencies iteratively. To do so, we used the theory of compressed sensing, adapted for handling correlated noise, and went through the following steps.

(i) Selecting a family of normalized vectors where the signal is represented by a small number of coefficients.

(ii) Approximating a solution to equation (9); for example, by discretizing the dictionary, and ensuring that the grid spacing is consistent with the noise power (see equation 15) then solving equation (11,$\mathrm{BP}_{\epsilon,W}$) with SPGL1 and take the average power. The introduction of the weight matrix $\mathbf{W}$ accounts for correlated Gaussian noises.

(iii) Estimating the detection significance, which we do by computing subsequent FAPs of the models with an increasing number of planets.

We showed that the published planets for each system could be seen directly on the same graph, and that taking into account the possible correlations in the noise could make a signal appear. This was established in the case of radial velocity data but the method could be adapted to other types of measurements, such as astrometric observations.

The use of the basis pursuit/ℓ_1-periodogram we suggest is as follows. This method can be used as a first guess to see if the signal is sparse or not; in that extent, it constitutes an evaluation of the

difficulty of the system and possibly a shortcut to the solution. It can bring attention to signal features that are hidden in the classical periodogram, which can still be used for an analysis 'by hand'. Secondly, for confirming the planetary nature of a system, we advocate to use in a second time statistical hypothesis testing.

The perspective for future work is twofold. First, we saw that the algorithm itself could be improved. Also, there might be significance tests more robust than the FAP, and the effect of introducing a weight matrix $\mathbf{W}$ must be studied into more depth. Secondly, let us recall that our method uses a priori information, that is the sparsity of the signal, but still does not handle all the information we have. To improve the technique, we wish to broaden its field of application by

(i) adapting the method for very eccentric orbits, through the addition of Keplerian vectors to the dictionary for example;

(ii) using precise models of the noise, especially magnetic activity, granulation, p-modes. Possibly include an adaptive estimation of the noise, especially one could extend the dictionary to wavelets;

(iii) handling several types of measurements at once (e.g. radial velocity, astrometry and photometry).

ACKNOWLEDGEMENTS

The authors wish to thank the anonymous referee for his insightful suggestions. NH thanks Evgeni Grishin for pointing out the matched filter technique to him. AC acknowledges support from CIDMA strategic project UID/MAT/04106/2013.

REFERENCES

Aigrain S., Gibson N., Roberts S., Evans T., McQuillan A., Reece S., Osborne M., 2011, in AAS/Division for Extreme Solar Systems Abstracts, Vol. 2, p. 11.05
Anglada-Escudé G., Tuomi M., 2012, A&A, 548, A58
Anglada-Escudé G., López-Morales M., Chambers J. E., 2010, ApJ, 709, 168
Arildsen T., Larsen T., 2014, Signal Process., 98, 275
Babu P., Stoica P., 2010, Digit. Signal Process., 20, 359
Babu P., Stoica P., Li J., Chen Z., Ge J., 2010, AJ, 139, 783
Baluev R. V., 2008, MNRAS, 385, 1279
Baluev R. V., 2009, MNRAS, 393, 969
Baluev R. V., 2011, Celest. Mech. Dyn. Astron., 111, 235
Baluev R. V., 2013a, Astron. Comput., 3, 50
Baluev R. V., 2013b, MNRAS, 436, 807
Baluev R. V., 2015a, MNRAS, 446, 1478
Baluev R. V., 2015b, MNRAS, 446, 1493
Becker S., Bobin J., Candès E. J., 2011, SIAM J. Imaging Sci., 4, 1
Bellmann K., 1975, Biometr. Z., 17, 271
Bobin J., Starck J.-L., Ottensamer R., 2008, IEEE J. Sel. Top. Signal Process., 2, 718
Boisse I. et al., 2009, A&A, 495, 959
Bourguignon S., Carfantan H., Böhm T., 2007, A&A, 462, 379
Butler R. P., Marcy G. W., Williams E., Hauser H., Shirts P., 1997, ApJ, 474, L115
Candès E., Fernandez-Granda C., 2013, J. Fourier Anal. Appl., 19, 1229
Candès E., Fernandez-Granda C., 2014, Commun., Pure Appl. Math., 67, 906
Candès E., Romberg J., Tao T., 2006a, IEEE Trans. Inf. Theory, 52, 489
Candès E. J., Romberg J. K., Tao T., 2006b, Commun. Pure Appl. Math., 59, 1207
Chandrasekaran V., Recht B., Parrilo P. A., Willsky A. S., 2010, preprint (arXiv:1012.0621)
Chen Y., Chi Y., 2014, IEEE Trans. Inf. Theory, 60, 6576
Chen S. S., Donoho D. L., Saunders M. A., 1998, SIAM J. Sci. Comput., 20, 33
Cohen A., Dahmen W., Devore R., 2009, J. Am. Math. Soc., 22, 211
Correia A. C. M. et al., 2010, A&A, 511, A21
Cumming A., 2004, MNRAS, 354, 1165
Cumming A., Marcy G. W., Butler R. P., 1999, ApJ, 526, 890
Daubechies I., DeVore R., Fornasier M., Güntürk C. S., 2010, Commun. Pure Appl. Math., 63, 1
Dawson R. I., Fabrycky D. C., 2010, ApJ, 722, 937
Delfosse X., Forveille T., Mayor M., Perrier C., Naef D., Queloz D., 1998, A&A, 338, L67
Demory B.-O. et al., 2011, A&A, 533, A114
Díaz R. F. et al., 2016, A&A, 585, A134
Donoho D., 2006, IEEE Trans. Inf. Theory, 52, 1289
Donoho D. L., Elad M., Temlyakov V. N., 2006, IEEE Trans. Inf. Theory, 52, 6
Duarte M. F., Baraniuk R. G., 2013, Appl. Comput. Harmon. Anal., 35, 111
Dumusque X., 2016, A&A, 593, A5
Dumusque X., Boisse I., Santos N. C., 2014, ApJ, 796, 132
Dumusque X. et al., 2016, preprint (arXiv:1609.03674)
Endl M. et al., 2012, ApJ, 759, 19
Engelbrecht C. A., 2013, in Guzik J. A., Chaplin W. J., Handler G., Pigulski A., eds, Proc. IAU Symp. 301, Precision Asteroseismology. Cambridge Univ. Press, Cambridge, p. 77
Ferraz-Mello S., 1981, AJ, 86, 619
Fischer D. A. et al., 2008, ApJ, 675, 790
Foster G., 1995, AJ, 109, 1889
Ge D., Jiang X., Ye Y., 2011, Math. Program., 129, 285
Gorodnitsky I. F., Rao B. D., 1997, IEEE Trans. Signal Process., 45, 600
Grant M., Boyd S., 2008, in Blondel V., Boyd S., Kimura H., eds, Lecture Notes in Control and Information Sciences, Recent Advances in Learning and Control. Springer-Verlag, Berlin, p. 95
Gregory P. C., 2011, MNRAS, 410, 94
Gregory P. C., 2016, MNRAS, 458, 2604
Horne J. H., Baliunas S. L., 1986, ApJ, 302, 757
Jenkins J. S., Yoma N. B., Rojo P., Mahu R., Wuth J., 2014, MNRAS, 441, 2253
Kay S. M., 1993, Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice-Hall, Englewood Cliffs, NJ
Kay S., Marple S. L. J., 1981, Proc. IEEE, 69, 1380
Laskar J., 1988, A&A, 198, 341
Laskar J., 1993, Celest. Mech. Dyn. Astron., 56, 191
Laskar J., 2003, preprint (arXiv:math/0305364)
Laskar J., Froeschlé C., Celletti A., 1992, Physica D, 56, 253
Laskar J., Boué G., Correia A. C. M., 2012, A&A, 538, A105
Lomb N. R., 1976, Ap&SS, 39, 447
Lovis C. et al., 2006, Nature, 441, 305
Lovis C. et al., 2011, A&A, 528, A112
McArthur B. E. et al., 2004, ApJ, 614, L81
Mallat S. G., Zhang Z., 1993, IEEE Trans. Signal Process, 41, 3397.
Marcy G. W., Butler R. P., Vogt S. S., Fischer D., Lissauer J. J., 1998, ApJ, 505, L147
Marcy G. W., Butler R. P., Fischer D. A., Laughlin G., Vogt S. S., Henry G. W., Pourbaix D., 2002, ApJ, 581, 1375
Melo C. et al., 2007, A&A, 467, 721
Meunier N., Lagrange A.-M., De Bondt K., 2012, A&A, 545, A87
Mishali M., Eldar Y., Tropp J., 2008, in IEEE 25th Convention of Electrical and Electronics Engineers in Israel, IEEEI 2008. p. 290
Mortier A., Faria J. P., Correia C. M., Santerne A., Santos N. C., 2015, A&A, 573, A101
Nelson B. E., Ford E. B., Wright J. T., Fischer D. A., von Braun K., Howard A. W., Payne M. J., Dindar S., 2014, MNRAS, 441, 442
Nelson B. E., Robertson P. M., Payne M. J., Pritchard S. M., Deck K. M., Ford E. B., Wright J. T., Isaacson H. T., 2016, MNRAS, 455, 2484
O'Toole S. J., Tinney C. G., Jones H. R. A., Butler R. P., Marcy G. W., Carter B., Bailey J., 2009, MNRAS, 392, 641
Pati Y. C., Rezaiifar R., Krishnaprasad P. S., 1993, in Proc. of the 27th Asilomar Conference on Signals, Systems and Computers, p. 40
Pelat D., 2013, Bases et Méthodes pour le Traitement de Données. Observatoire de Paris

Queloz D. et al., 2001, A&A, 379, 279
Rajpaul V., Aigrain S., Osborne M. A., Reece S., Roberts S. J., 2015, MNRAS, 452, 2269
Reegen P., 2007, A&A, 467, 1353
Rivera E. J. et al., 2005, ApJ, 634, 625
Rivera E. J., Laughlin G., Butler R. P., Vogt S. S., Haghighipour N., Meschiari S., 2010, ApJ, 719, 890
Roberts D. H., Lehar J., Dreher J. W., 1987, AJ, 93, 968
Rockafellar R. T., 1970, Convex Analysis. Princeton Univ. Press, Princeton, NJ
Scargle J. D., 1982, ApJ, 263, 835
Schuster A., 1898, Terr. Magn., 3, 13
Schwarzenberg-Czerny A., 1998, Balt. Astron., 7, 43
Ségransan D. et al., 2011, A&A, 535, A54
Starck J.-L., Elad M., Donoho D. L., 2005, IEEE Trans. Image Process., 14, 1570
Stoica P., Babu P., 2012, Signal Process., 92, 1580
Sulis S., Mary D., Bigot L., 2016, preprint (arXiv:1601.07375)
Tang G., Bhaskar B., Recht B., 2013a, in 2013 Asilomar Conference on Signals, Systems and Computers, p. 1043
Tang G., Bhaskar B., Shah P., Recht B., 2013b, IEEE Trans. Inf. Theory, 59, 7465
Tibshirani R., 1994, J. R. Stat. Soc. Ser. B, 58, 267
Tropp J. A., Gilbert A. C., 2007, IEEE Trans. Inf. Theory, 53, 4655
Tropp J. A., Laska J. N., Duarte M. F., Romberg J. K., Baraniuk R. G., 2010, IEEE Trans. Inf. Theory, 56, 520
Tuomi M., 2012, A&A, 543, A52
Tuomi M. et al., 2013, A&A, 551, A79
Tuomi M., Jones H. R. A., Barnes J. R., Anglada-Escudé G., Jenkins J. S., 2014, MNRAS, 441, 1545
van den Berg E., Friedlander M. P., 2008, SIAM J. Sci. Comput., 31, 890
Winn J. N. et al., 2011, ApJ, 737, L18
Wisdom J., 2005, BAAS, 37, 525
Zechmeister M., Kürster M., 2009, A&A, 496, 577
Zucker S., 2015, MNRAS, 449, 2723
Zucker S., 2016, MNRAS, 457, L118

APPENDIX A: MINIMUM GRID SPACING

Let us consider a signal made of p pure harmonics sampled at times $\boldsymbol{t} = (t_k)_{k=1,\dots,m}$, $\boldsymbol{y} = \sum_{j=1}^{p} c_j \mathrm{e}^{\mathrm{i}\omega_j t}$. We denote by ω'_j and $\Delta\omega$ two real numbers such that for each j

$$\Delta\omega < \frac{4}{T} \tag{A1}$$

$$|\omega_j - \omega'_j| < \Delta\omega\ , \tag{A2}$$

where $T = t_m - t_1$. For each t_k and each j,

$$|c_j||e^{\mathrm{i}\omega_j t_k} - e^{\mathrm{i}\omega'_j t_k}| = |c_j| \left| \mathrm{e}^{\mathrm{i}\frac{\omega_j+\omega'_j}{2}t_k} \left(\mathrm{e}^{\mathrm{i}\frac{\omega_j-\omega'_j}{2}t_k} - \mathrm{e}^{-\mathrm{i}\frac{\omega_j-\omega'_j}{2}t_k} \right) \right|$$
$$= 2|c_j| \left| \sin\left(\frac{\omega_j - \omega'_j}{2} t_k \right) \right|.$$

So denoting $\boldsymbol{y}' = \sum_{j=1}^{p} c_j \mathrm{e}^{\mathrm{i}\omega'_j \boldsymbol{t}}$,

$$|y_k - y'_k| = \left| \sum_{j=1}^{p} c_j \left(\mathrm{e}^{\mathrm{i}\omega_j t_k} - \mathrm{e}^{\mathrm{i}\omega'_j t_k} \right) \right|$$
$$\leqslant 2 \sum_{j=1}^{p} \left| c_j \sin\left(\frac{\omega_j - \omega'_j}{2} t_k \right) \right|.$$

Without loss of generality, the origin of time is shifted to $-T/2$; therefore,

$$2 \sum_{j=1}^{p} \left| c_j \sin\left(\frac{\omega_j - \omega'_j}{2} t_k \right) \right| \leqslant \sin \frac{\Delta\omega T}{4} \sqrt{\sum_{j=1}^{p} |c_j|^2}. \tag{A3}$$

Finally, a condition for y' to be an acceptable solution is

$$\|\mathbf{W}(y - \boldsymbol{y}')\|_{\ell_2}^2 \leqslant \|\mathbf{W}\|^2 \|\boldsymbol{y} - \boldsymbol{y}'\|_{\ell_2}^2$$
$$\leqslant \|\mathbf{W}\|^2 \sum_{k=1}^{m} |y_k - y'_k|^2$$
$$\leqslant 4\|\mathbf{W}\|^2 \sum_{k=1}^{m} \left(\sum_{j=1}^{p} \left| c_j \sin\left(\frac{\omega_j - \omega'_j}{2} t_k \right) \right| \right)^2$$

given (equation A3),

$$\leqslant 4m\|\mathbf{W}\|^2 \sin^2 \frac{\Delta\omega T}{4} \sum_{j=1}^{p} |c_j|^2$$

where $\|\mathbf{W}\| = \sup_{x \in \mathbb{C}^m} \frac{\|\mathbf{W}x\|_{\ell_2}}{\|x\|_{\ell_2}}$. When the matrix $\mathbf{W}$ is diagonal, the formula can be improved:

$$\|\mathbf{W}(\boldsymbol{y} - \boldsymbol{y}')\|_{\ell_2}^2 = \sum_{k=1}^{m} \frac{|y_k - y'_k|^2}{\sigma_k^2}$$
$$\leqslant 4 \sum_{k=1}^{m} \frac{1}{\sigma_k^2} \left(\sum_{j=1}^{p} \left| c_j \sin\left(\frac{\omega_j - \omega'_j}{2} t_k \right) \right| \right)^2$$

given (equation A3),

$$\leqslant 4 \sin^2 \frac{\Delta\omega T}{4} \sum_{j=1}^{p} |c_j|^2 \sum_{k=1}^{m} \frac{1}{\sigma_k^2}.$$

So ϵ_{grid} can be chosen as

$$\epsilon_{\mathrm{grid}} = 2 \sqrt{\sum_{j=1}^{p} |c_j|^2} \sqrt{\sum_{k=1}^{m} \frac{1}{\sigma_k^2}} \sin \frac{\Delta\omega T}{4}. \tag{A4}$$

And conversely given an ϵ, the grid spacing that ensures that there exists a vector that has the correct ℓ_0 norm is

$$\Delta\omega = \frac{4}{T} \arcsin \frac{\epsilon}{2 \sqrt{\sum_{j=1}^{p} |c_j|^2} \sqrt{\sum_{k=1}^{m} \frac{1}{\sigma_k^2}}}. \tag{A5}$$

APPENDIX B: DIGGING IN RED NOISE WITH NON-DIAGONAL *W*

B1 Short period buried in the noise

Our method uses the tools of compressed sensing, especially the algorithms to minimize ℓ_1 norms with the constraint that the reconstructed signal is not too far from the observations [see equation (5)]. To the best of our knowledge, the case where the noise is correlated has been considered only in Arildsen & Larsen (2014), and is not specialized for Gaussian processes. Here, we introduce a weight matrix and obtain problem (11,BP$_{\epsilon,W}$), reproduced here:

$$x^\star = \arg\min_{x \in \mathbb{C}^n} \|\boldsymbol{x}\|_{\ell_1} \quad \text{s. t.} \quad \|\mathbf{W}(Ax - \boldsymbol{y})\|_{\ell_2} \leqslant \epsilon. \tag{BP$_{\epsilon,\mathbf{W}}$}$$

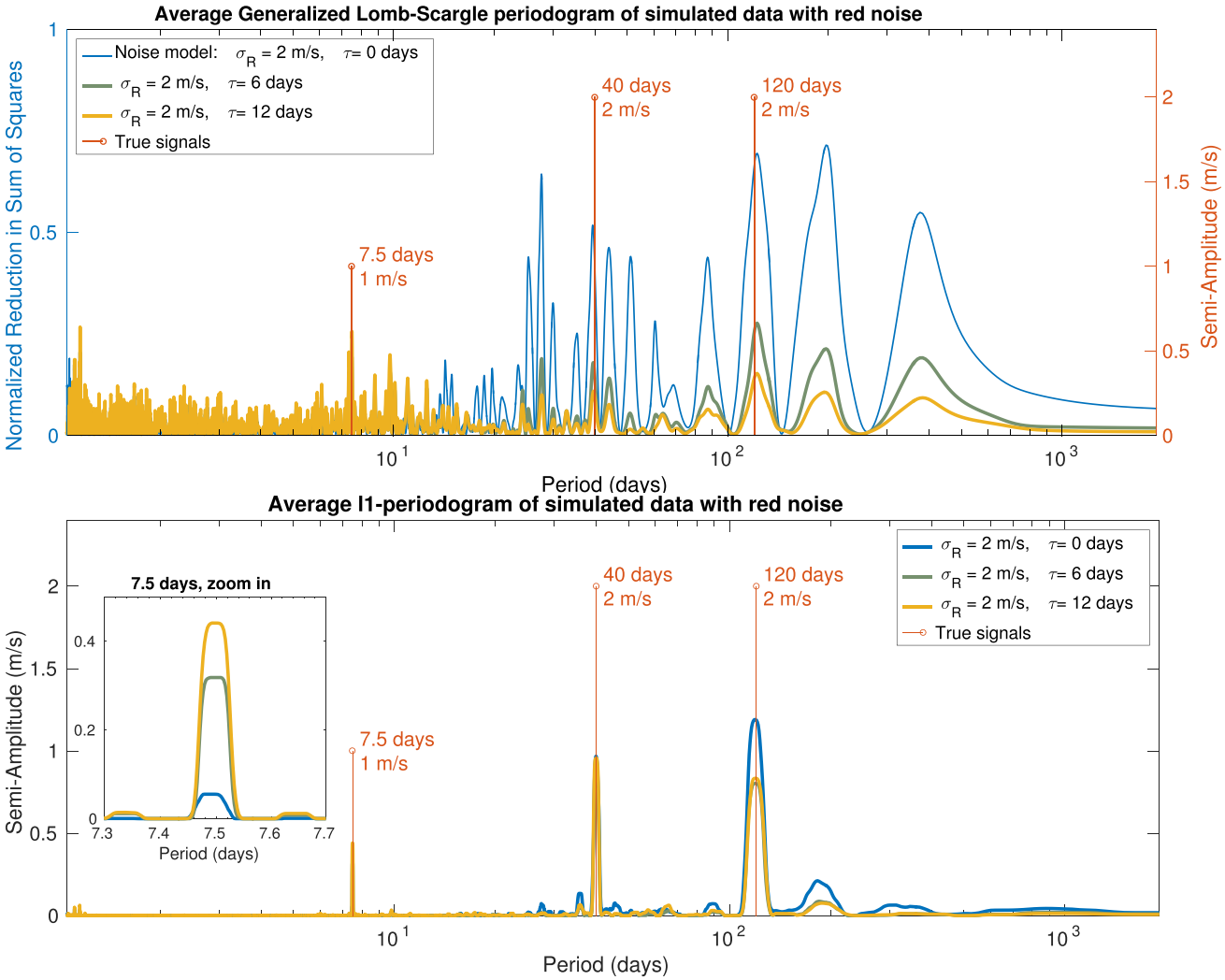

Figure B1. Average ℓ_1-periodogram for 50 data sets generated with red noise of characteristics $\sigma_W = 0$, $\sigma_R = 2\ \mathrm{m\ s^{-1}}$ and $\tau = 12$ d according to model (14). The curves correspond to the solutions of equation (11,BP$_{\epsilon,W}$) with different weight matrices W whose parameters are $\sigma_W = 0$, $\sigma_R = 2\ \mathrm{m\ s^{-1}}$ and $\tau = 0$, 6 or 12 d(respectively the blue, green and yellow curves).

To illustrate the interest of choosing an appropriate weight matrix, we will show an example where acknowledging the red noise makes a planet visible. Let us first consider a data set constructed as follows.

(i) The measurement times are those of HD 69830 (74 measurements).

(ii) The true signal is $y(t) = 1\cos(\frac{2\pi}{7.5}t) + 2\cos(\frac{2\pi}{40}t + 2) + 2\cos(\frac{2\pi}{120}t + 1)\ \mathrm{m\ s^{-1}}$.

(iii) The noise is red, with parameters $\sigma_W = 0$, $\sigma_R = 2\ \mathrm{m\ s^{-1}}$ and $\tau = 12$ d, where $\sigma_{\mathbf{W}}$, σ_R and τ are the parameters of the autocorrelation function R defined in equation (14) reproduced here:

$$R(\Delta t) = \sigma_R^2 \mathrm{e}^{-\frac{|\Delta t|}{\tau}}, \quad \Delta t \neq 0$$

$$R(0) = \sigma_W^2 + \sigma_R^2.$$

The noise defined above is such that its correlation with low frequencies is higher than with high frequencies.

We test if changing the weight matrix could allow us to find signals that would not be seen otherwise. To do so, 50 noise time series $(n_k(\boldsymbol{t}))_{k=1,\ldots,50}$ are generated and the method is applied to each $y_k(\boldsymbol{t}) = \boldsymbol{y}(\boldsymbol{t}) + n_k(\boldsymbol{t})$ for three different weight matrices, all other parameters being fixed. In each case, they are defined according to model (14) with $\sigma_W = 0$, $\sigma_R = 2\ \mathrm{m\ s^{-1}}$ and $\tau = 0$, 6 or 12 d. The grid goes between 0 and 0.95 cycles d^{-1} and ϵ verifies $F_{\chi^2_m}(\epsilon^2_{\mathrm{noise}}) = 0.1$. The resulting ℓ_1-periodograms are averaged (see Fig. B1b).

To compare with a classical approach, we also compute classical periodograms for the same signals $y_k(\boldsymbol{t})$ and average them. For the comparison to be fair, we fit the model parameters A, B, C in $A\cos\omega t + B\sin\omega t + C$ to $\boldsymbol{y}(\boldsymbol{t})$ with the same weight matrices as the ones used above. This gives Fig. B1(a). If the weight matrix is left diagonal, then the low-frequency terms dominate. Using the appropriate noise model gradually reduces the spurious low frequencies.

We stress two features: as the noise model becomes accurate, the short period becomes apparent, which justifies the trial of different noise matrices on real radial velocity data sets to see if a peak appears. Secondly, when $\mathbf{W}$ is defined with an exponential autocorrelation function, the estimation of the peaks becomes biased: some frequencies will have a tendency to be interpreted by the algorithm as noise. The amplitude of the 120 d periodicity is then underestimated. This bias could prevent us from finding small amplitudes when using non-diagonal weight matrices. When the number of frequency in the signal increases, the bias becomes more

complicated. In order to mitigate this effect, we suggest to decrease the value of ϵ when testing different noise model. Thus, the model 'sticks' to the observations, and if a periodicity truly is in the data, the chance of it being too underestimated decreases. This is why, we took $\epsilon_{\rm noise}$ such that $F_{\chi^2_m}(\epsilon^2_{\rm noise}) = 0.1$ and not $F_{\chi^2_m}(\epsilon^2_{\rm noise}) = 0.999$, which would reject more signals in the residual.

B2 No automatic procedure so far

Here the improvement due to an appropriate handle of the noise is seen by eye. One could wonder if a simple criterion could allow one to chose an appropriate weight matrix automatically. In all cases when the algorithm has converged, we have $\|\mathbf{W}(\mathbf{A}\boldsymbol{x} - \boldsymbol{y})\|_{\ell_2} = \epsilon$ to a certain tolerance, or $x = 0$. Looking at the χ^2 of the residuals as usual is then not appropriate.

As in all cases the columns of matrix $\mathbf{W}A$ and the weighted observations Wy are normalized. Therefore, the problem always comes down to minimizing

$$x^\star = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{x}\|_{\ell_1} \text{ s. t. } \|\mathbf{A}'\boldsymbol{x} - \boldsymbol{y}'\|_{\ell_2} \leqslant \epsilon, \tag{B1}$$

where $\mathbf{A}'$ has normed columns and $\boldsymbol{y}'$ is a unitary vector. It is then tempting to see if there is a correlation between the ℓ_0 or ℓ_1 norm of $x^\star$ and the success of the method. Unfortunately, this is not the case. Whether there is an automatic way to select the appropriate weight matrix remains an open question.

APPENDIX C: SPURIOUS TALLEST PEAK OF THE GLS PERIODOGRAM

In this section, we show examples where the initial highest peak of the periodogram is spurious due to aliasing. We take the 74 measurement dates of HD 69830 and generate 500 systems with three circular orbits with the following properties.

(i) The amplitudes are those of the three Neptunes of HD 69830 (2.2, 2.66 and 3.51 m s^{-1}).
(ii) The periods P_1, P_2, P_3, are selected uniformly in $\log P$ in the range 1.2–2000 d.
(iii) The phases are uniformly distributed on $[0, 2\pi]$.
(iv) The noise standard deviation is 0.6 m s^{-1}.

We compute the number of times the maximum peaks of the GLS and ℓ_1-periodogram are spurious. The criterion we take for failure is when the frequency of the highest peak and any of the three true frequencies are greater than the inverse of the total observation time, that is $|1/P_{1,2,3} - 1/P_{\rm max}| > 1/T_{\rm obs}$.

Fig. C1 shows the GLS periodogram and ℓ_1-periodogram of representative cases where the highest peak of the GLS periodogram is spurious. In these conditions, when searching for periods in the 1.2–2000 d with the periodogram, we find that the strongest peak is spurious in 33 cases out of 500 simulations, while the tallest peak of the ℓ_1-periodogram only was incorrect in two cases. In those, the GLS periodogram was also failing.

An interesting feature of the cases where the ℓ_1-periodogram fails is that one can see that the solution is not sparse. This is a very useful property we observed empirically: we have not found any occurrence of ℓ_1-periodogram that looks clean, with well-separated clear peaks, where one of the peaks was completely spurious. We display one of the two failures of the ℓ_1-periodogram in Fig. C2. First of all neither the GLS nor the ℓ_1-periodogram leads the observer completely astray. Secondly, we see that as opposed to the ℓ_1-periodogram of the systems studied here, the figure is not clean, which should invite the analyst to a certain suspicion.

APPENDIX D: FITTING THE ANCILLARY MEASUREMENTS

In Section 4.6, we suggest to fit the activity indicators to the radial velocity time series. The present discussion wishes to give a justification to this approach. The idea is to exploit the possible correlations between radial velocity and ancillary measurements when the star is active. For instance, on the first system of the RV Fitting Challenge (Dumusque et al. 2016) where activity dominates the signal, the radial velocity, FWHM, bisector span and $\log R'_{HK}$ exhibit very similar features at low frequency (see Fig. D1).

Let us approximate the error made when fitting an ancillary indicator. We consider the radial velocity signal $\boldsymbol{y}(t) = \boldsymbol{P}(t) + \boldsymbol{a}(t) + \boldsymbol{\epsilon}(t)$, where $\boldsymbol{P}(t)$ is due to a planetary companion, $\boldsymbol{y}(a)$ is a deterministic signal due to activity and $\boldsymbol{\epsilon}$ is a Gaussian noise of covariance matrix $\mathbf{V}$. We also consider an ancillary measurement $\boldsymbol{z}(t) = \boldsymbol{a}(t) + \boldsymbol{\epsilon}'$, where $\boldsymbol{\epsilon}'(t)$ is another Gaussian noise of covariance matrix $\mathbf{V}$. If we fit $\boldsymbol{z}(t)$ to $\boldsymbol{y}(t)$, we obtain (dropping the t notation)

$$\boldsymbol{y}_{\rm detrend} = \boldsymbol{y} - \boldsymbol{y}_{\rm fit} = \boldsymbol{y} - \frac{\boldsymbol{z}^T\mathbf{V}^{-1}\boldsymbol{y}}{\boldsymbol{z}^T\mathbf{V}^{-1}\boldsymbol{z}}\boldsymbol{z} \tag{D1}$$

$$\boldsymbol{y}_{\rm detrend} = \boldsymbol{y} - \frac{(\boldsymbol{a}+\boldsymbol{\epsilon}')^T\mathbf{V}^{-1}(\boldsymbol{P}+\boldsymbol{a}+\boldsymbol{\epsilon})^T}{(\boldsymbol{a}+\boldsymbol{\epsilon}')^T\mathbf{V}^{-1}(\boldsymbol{a}+\boldsymbol{\epsilon})}(\boldsymbol{a}+\boldsymbol{\epsilon}'). \tag{D2}$$

We assume that the noise is small compared to $\boldsymbol{a}$, which allows us to develop the denominator at first order in $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}'$

$$y_{\rm fit} \approx \frac{(\boldsymbol{a}+\boldsymbol{\epsilon}')^T\mathbf{V}^{-1}(\boldsymbol{P}+\boldsymbol{a}+\boldsymbol{\epsilon})}{\boldsymbol{a}^T\mathbf{V}^{-1}\boldsymbol{a}} \times \left(1 - \frac{\boldsymbol{\epsilon}'^T\mathbf{V}^{-1}\boldsymbol{a}}{\boldsymbol{a}^T\mathbf{V}^{-1}\boldsymbol{a}} - \frac{\boldsymbol{\epsilon}^T\mathbf{V}^{-1}\boldsymbol{a}}{\boldsymbol{a}^T\mathbf{V}^{-1}\boldsymbol{a}}\right)(\boldsymbol{a}+\boldsymbol{\epsilon}').$$

After developing that expression at first order in $\boldsymbol{\epsilon}$ and ϵ', we compute its mathematical expectancy taking into account only the zero-order, ϵ^2 and ϵ'^2 coefficients. In the simple case where the noise is i.i.d. of variance σ^2, we obtain

$$\mathbb{E}\{\boldsymbol{y}_{\rm fit}\} \approx \frac{\sigma^2}{\|\boldsymbol{a}\|^2}\boldsymbol{P} \tag{D3}$$

$$+ \left(1 + \frac{\boldsymbol{a}^T\boldsymbol{P}}{\|\boldsymbol{a}\|^2_{\ell_2}} - \frac{2\sigma^2}{\|\boldsymbol{a}\|^2_{\ell_2}} - \frac{\|\boldsymbol{P}\|_{\ell_2}\sigma^2}{\|\boldsymbol{a}\|^3_{\ell_2}} - \frac{\boldsymbol{a}^T\boldsymbol{P}\sigma^2}{\|\boldsymbol{a}\|^4_{\ell_2}}\right)\boldsymbol{a}. \tag{D4}$$

We would like $\boldsymbol{y}_{\rm fit}$ to be as close to $\boldsymbol{a}$ as possible. This will be better satisfied as the correlation $\boldsymbol{a}^T\boldsymbol{P}$ and the signal-to-noise $\sigma^2/\|\mathbf{a}\|_{\ell_2}$ decrease. The fact that a term $\boldsymbol{a}^T\boldsymbol{P}$ appears in the equation above should not be surprising. The mutual coherence defined in Section 5.2 grasps that the correlation between the parts of the model is an obstacle to the recovery of the true signals.

For the RV Fitting Challenge, not only have we fitted one activity indicator but several. We point out that this approach is consistent with Rajpaul et al. (2015). Indeed, they consider that the activity-induced variations of the measurements depend linearly on an underlying zero-mean Gaussian process $G(t) = F^2(t)$ and its derivative $\dot{G}(t)$, where $F(t)$ is the fraction of the sphere covered with spots. The evolution of the indicators is modelled by formulae (14)–(16), reproduced below,

$$\Delta{\rm RV} = V_c G(t) + V_r \dot{G}(t), \tag{D5}$$

$$\log R'_{HK} = L_c G(t), \tag{D6}$$

$${\rm BIS} = B_c G(t) + B_r \dot{G}(t), \tag{D7}$$

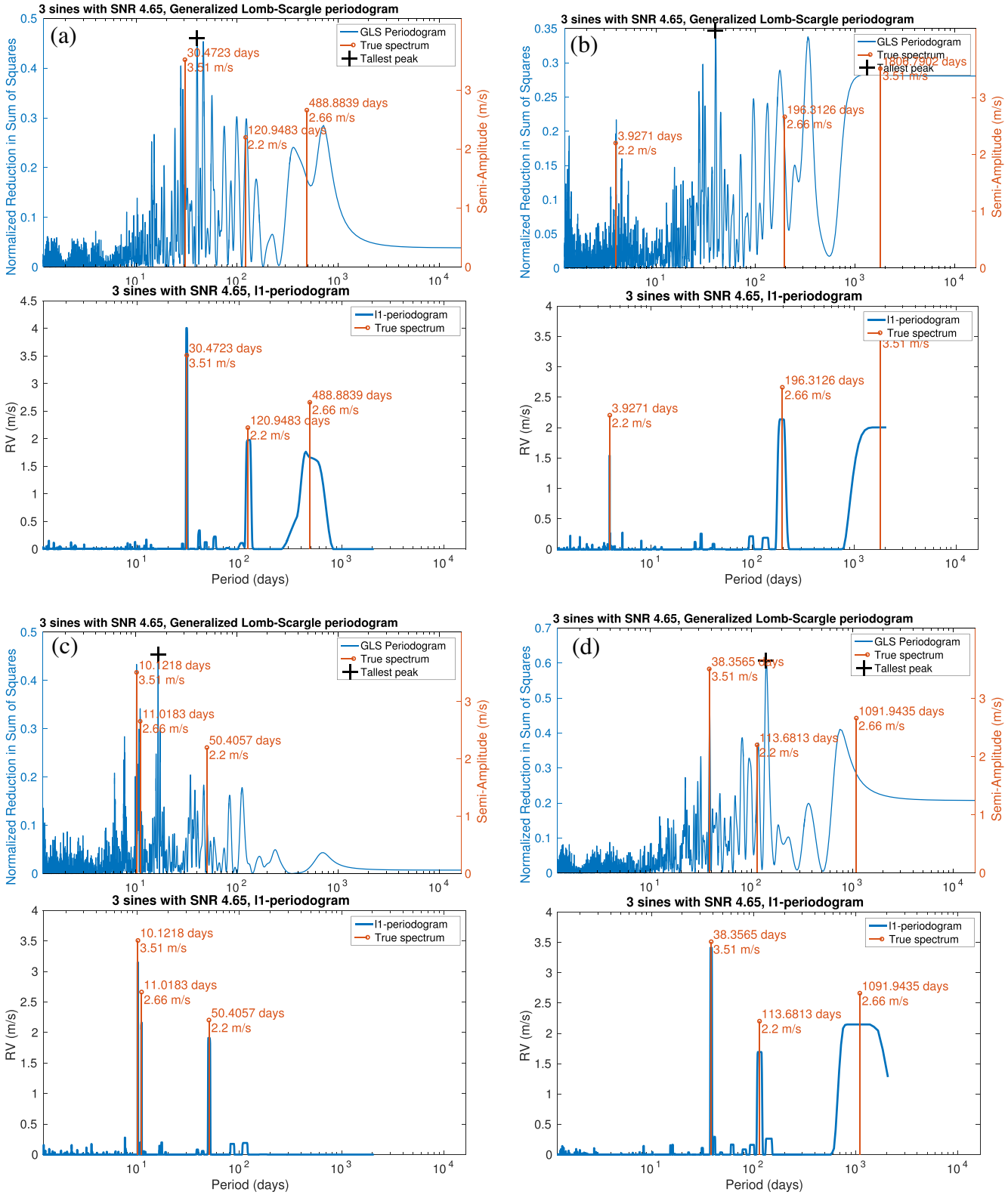

Figure C1. Peak amplitudes and associated FAPs for the four systems analysed.

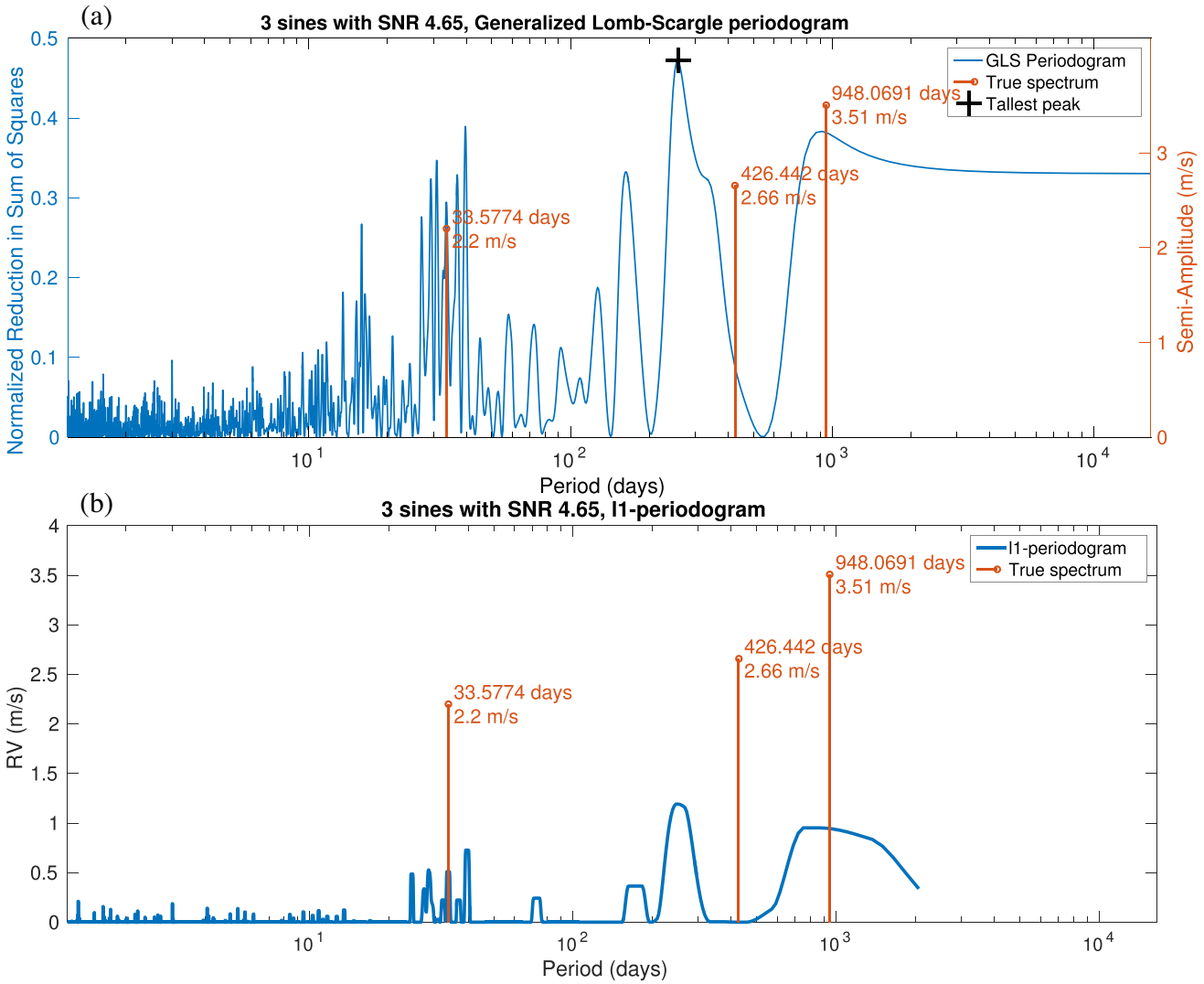

Figure C2. Failure of the GLS (a) and ℓ_1 (b) periodograms.

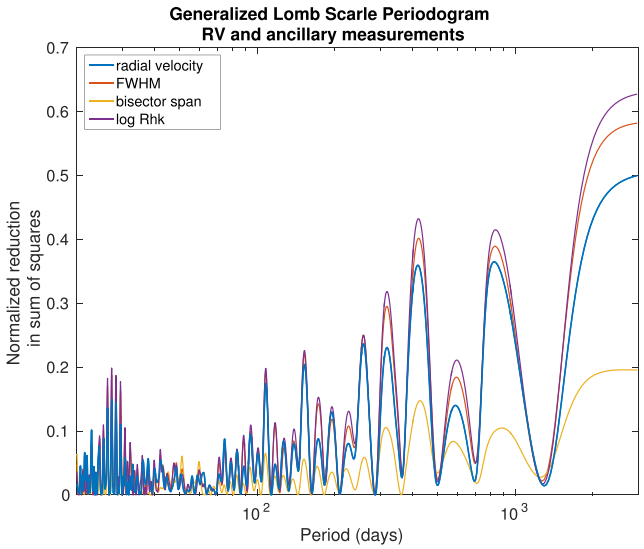

Figure D1. GLS periodogram of radial velocity and ancillary measurements at low frequencies.

for some constants V_c, V_r, L_c, B_c, B_r. This means that for a given realization $(\boldsymbol{g}, \boldsymbol{g}')$ of $(G(t), \dot{G}(t))$, the subspace generated by the log R'_{HK} and the bisector span BIS is the same as the space generated by $\boldsymbol{g}, \boldsymbol{g}'$. So according to that model, projecting the radial velocity on to (log R'_{HK},BIS) is equivalent to projecting on to $(\boldsymbol{g}, \boldsymbol{g}')$.

However, there is an uncertainty on the behaviour of the ancillary measurements and additional noise. We have to decide if fitting an uncertain model is better than working with the raw data. One thing that could happen is that fitting the combination of the three ancillary measurements would greatly change the spectral content of the radial velocity time series by absorbing some frequencies, potentially due to planets. To estimate this risk, we first compute the term $\boldsymbol{a}^T \boldsymbol{P}/\|\boldsymbol{a}\|^2_{\ell_2}$ in equation (D4), assuming that the signal $\boldsymbol{y} = \boldsymbol{P} = \mathrm{e}^{\mathrm{i}\omega t}$ is a pure harmonic of amplitude 1 m s^{-1}. Here $\boldsymbol{a}$ designates the FWHM, bisector span or log R'_{HK}, respectively, shown by the red, yellow and purple curves in Fig. D2. We also compute the fraction of the energy of the signal before and after the fit of the three ancillary measurements simultaneously, that is

$$\text{Fraction}(\omega) = \frac{(\boldsymbol{y}_\omega - \boldsymbol{y}_{\text{fit}})^T \mathbf{V}^{-1} (\boldsymbol{y}_\omega - \boldsymbol{y}_{\text{fit}})}{\boldsymbol{y}_\omega^T \mathbf{V}^{-1} \boldsymbol{y}_\omega}; \tag{D8}$$

this one is represented by the blue curve in Fig. D2.

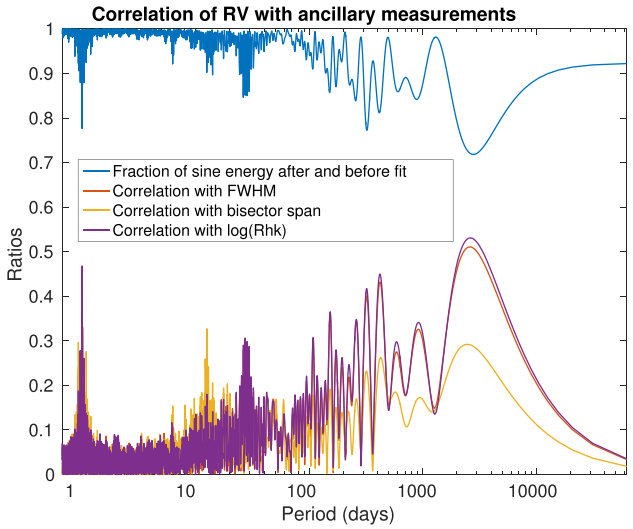

Figure D2. Energy of a cosine function after the fit of the FWHM, bisector span, log R'_{HK} and a constant.

For the system analysed in Section 4.6, only 15 per cent of the energy is absorbed in general, with a maximum of 27 per cent at a period of 2000 d. The peaks at 25 and 12.5 d correspond to the rotation period of the star and its first harmonic, which are expected to be correlated with the radial velocity and ancillary measurements.

This discussion does not intend to provide strong statistical arguments, but rather to show that the spectral content should not be too affected by fitting the FWHM, bisector span and log R'_{HK}.

This paper has been typeset from a TEX/LATEX file prepared by the author.