Radial Velocity Data Analysis with Compressed Sensing Techniques

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

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

7 September 2016

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 processes 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: Radial Velocity – Sparse Recovery – Orbit Estimation

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 from finding small planets, as pointed out for instance by Anglada-Escudé et al. (2010); Tuomi (2012). There might even be 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 suboptimal solution, random searches such as Monte Carlo Markov Chain (MCMC) 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 et al. 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 an *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 assumptions, stable planetary systems are quasi-periodic as well (e.g. Laskar 1993). By doing so, the periodogram can be efficiently cleaned (see figures 1,2,3,4,5).

The field of signal processing devoted to the study of sparse signals is often referred to as "Compressed Sensing" or "Compressive Sampling" (Donoho 2006; Candès et al. 2006b) – 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 to recover exactly a spectrum while sampling it at a much lower rate than the Nyquist frequency (Mishali et al. 2008; Tropp et al. 2009). Its use was advocated to improve the scientific data transmission in space-based astronomy (Bobin et al. 2008). Sparse recovery techniques are also used in image processing (e.g. Starck et al. 2005).

It seems relevant to add to that list a few techniques developed by astronomers to retrieve harmonics in a signal. In

[*] E-mail:nathan.hara@obspm.fr

the next section, we show that even though the term "sparsity" is not explicitly used (except in Bourguignon et al. 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 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 the method is discussed section 5 and conclusions are drawn 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 generalisation 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 to 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 et al. 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 et al. 1993; Tropp & Gilbert 2007). In that case, when an 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 et al. 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 peri-

odogram", 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 et al. 1998; Starck et al. 2005) and iteratively re-weighted least squares (IRWLS) (e.g. Gorodnitsky & Rao 1997; Donoho 2006; Candès et al. 2006a; Daubechies et al. 2010). In the context of astronomy, Bourguignon et al. (2007) implements a convex relaxation method using ℓ_1 norm weighting (see equation (2)) to find periodicity in unevenly sampled signals and Babu et al. (2010) presents 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

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

where $y(t)$ is a vector made of m measurements, and $x^\star = \arg\min f(x)$ denotes the element such that $f(x^\star) = \min f(x)$ 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 et al. (2009) in the noise-free case. Solving that problem is suggested by Baluev (2013b) under the name of "multi-frequency periodograms". However, finding that minimum by discretizing the values of $(K_j, \omega_j, \phi_j)_{j=1..k}$ depends exponentially on the number of parameters, and the multi-frequency periodograms could hardly handle more than three or four sines with conventional methods. However, with parallel progamming on GPUs one can handle up to ≈ 25 frequencies depending on the number of measurements (Baluev 2013a). Jenkins et al. (2014) explicitly mentions the above problem and suggests a tree-like algorithm to explore the frequency space. They analyse GJ 876 with their procedure and find six significant harmonics, which we confirm 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 are 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 $\Pr\{y(t)|\mathcal{M}_{k+1}\}/\Pr\{y(t)|\mathcal{M}_k\}$ is greater than 150 (see Tuomi et al. 2014), $y(t)$ 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 et al. (2006) shows that for a wide range of signals, this algorithm is outperformed by ℓ_1 relaxation methods. Does this claim still applies 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 appendix C that in some cases, the tallest peak of the periodogram is spurious but ℓ_1 minimization prevents from being mislead.

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 $\mathscr{A} = (a_j(t))_{j \in I}$ such that the signal to be analysed $y(t)$ is represented by a linear combination of a few elements of $\mathscr{A}$. 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 $a_{-\omega}(t) = e^{-i\omega t}$ and $a_\omega(t) = e^{i\omega t}$ where t is the array of measurement times.

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

- $y(t)$ denotes the vector of observations at times $t = t_1...t_m$, $y(t) \in \mathbb{R}^m$ for radial velocity data sets.
- The ℓ_p norm of a complex or real vector x with n components is defined as

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

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

- For a function f defined on a set E, $\arg\min_{x \in E} f(x)$ is the element for which the minimum is attained, that is $x^\star$ of E such that $f(x^\star) = \min_{x \in E} f(x)$. We denote by the superscript $\star$ the solution of the minimization problem under consideration. In all the cases considered here except (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 $(a_j(t))_{j=1...S}$ and their corresponding amplitudes x_j. To enhance the sparsity of the representation, one can think of solving

$$\arg\min_{\substack{a_j(t) \in \mathscr{A} \\ S \in \mathbb{C}}} S \quad \text{s.t.} \quad \left\| \sum_{j=1}^{S} x_j a_j(t) - y(t) \right\|_{\ell_2} \leqslant \varepsilon \tag{3}$$

that is finding the smallest number of elements of $\mathscr{A}$ required to approximate $y(t)$ with a certain tolerance ε. This one is *a priori* a combinatorial problem which seems unsolvable if $\mathscr{A}$ is infinite or of an exponential complexity if the dictionary is

finite. In the latter case $\mathscr{A}$ can be viewed as an $m \times n$ matrix A. In that case, on can re-write equation (3) like:

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

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

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

the problem becomes convex and still enhances sparsity efficiently. In the signal processing litterature, 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 (5) instead of (3). Let us cite a few results – among many: when $y(t)$ is noise free, Donoho (2006) shows that under certain hypotheses the solution to (5) is equal to the solution of (3); more generally, denoting by $y_t = Ax_t$ the true signal, such that $y = y_t + e$, e being the error, there is a theoretical bound on $\|Ax^\star - y_t\|_{\ell_2}$ (Candès et al. 2006b). One can also obtain constraints on $\|x^\star - x_t\|_{\ell_2}$ or conditions to have $\text{supp}(x^\star) \subset \text{supp}(x_t)_{\ell_2}$ where $\text{supp}(x)$ is the set of indices where x is non-zero (e.g. Donoho et al. 2006). In summary, there are results guaranteeing the performance for de-noising, compression and also for inverse problems, the search for planets being a particular case of the latter.

These results apply to a finite dictionary $\mathscr{A}$, but the periods of the planets could be anywhere: $\mathscr{A}$ 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 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 2012b). This one will give guidelines to improve our procedure.

2.2 Atomic norms minimization

If $\mathscr{A}$ is infinite, the ℓ_1 norm cannot be used straightforwardly. Chandrasekaran et al. (2010) suggests to use an "atomic norm" that extends (5) to infinite dictionaries. Practical methods to solve the new minimization problem are designed in Candès & Fernandez-Granda (2012a) and Tang et al. (2013b). The atomic norm $\|y\|_{\mathscr{A}}$, of $y \in \mathbb{R}^m$ or $\mathbb{C}^m$ defined for a dictionary $\mathscr{A}$ is the smallest ℓ_1 norm of a combination of vectors of the dictionary reproducing y:

$$\|y\|_{\mathscr{A}} = \inf \left\{ \sum_j |x_j|, y = \sum_j x_j a_j(t) \right\} \tag{6}$$

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

$$u^\star = \arg\min_{u \in \mathbb{C}^m} \|u - y(t)\|_{\ell_2}^2 + \lambda \|u\|_{\mathscr{A}} \tag{7}$$

where λ is a positive real number fixed according to the noise. This problem is often referred to as Atomic Norm De-Noising. 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 u. The quantities we are interested in are the dictionary elements $a_j^\star$

and the coefficients $x^\star$ selected by the minimization, where $u^\star = \sum_{j=1}^{S^\star} x_j^\star a_j^\star(t)$.

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 as Tuomi et al. (2013) or Rajpaul et al. (2015) stress that the detection efficiency and robustness improves as the noise model becomes more realistic. Aigrain et al. (2011) suggests to consider the RV time series as Gaussian processes: the noise $n(t)$ is then characterized by its covariance matrix V which is such that $V_{kl} = \mathbb{E}\{n(t_k)n(t_l)\}$, $\mathbb{E}$ 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 V is equivalent to choosing R. This approach is similar to Sulis et al. (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 $P(\omega) = |\mathscr{F}(R)|^2$ where $\mathscr{F}$ denotes the Fourier transform.

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

$$\ln(L) = -\frac{m}{2}\ln(2\pi) - \frac{1}{2}\det(V) - \frac{1}{2}(y - Ax)^T V^{-1}(y - Ax) \quad (8)$$

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

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

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 A diminishes (Donoho et al. 2006). This one is defined as the maximum correlation between two column-vectors of A. We here consider a weight matrix, but we can go back to the previous problem by noting that $W(Ax - y)$ can be re-written $A'x - y'$ where $A' = WA$ and $y' = Wy$. If we now consider two column vectors of A', $a_1' = Wa_1$ and $a_2' = Wa_2$, their correlation is $a_1'^T a_2' = a_1 W^T W a_2 = a_1 V^{-1} a_2$. In other words introducing a matrix 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 x in a signal $s = x + n$ where n is a noise of covariance matrix V if $xV^{-1}s \leqslant \gamma$ 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 V is diagonal and its elements are σ_k^2, where σ_k is the measurement error at time t_k. W is defined as $V^{-1/2}$ so is a diagonal matrix of elements $w_{kk} = 1/\sigma_k$. Therefore,

$a_1'^T a_2' = a_1 W^T W a_2 = \sum_{k=1}^{n} \frac{a_1(t_k)a_2(t_k)}{\sigma_k^2}$. 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 from having the benefits of a correct noise model is discussed in appendix B. We show that choosing an appropriate weight matrix W indeed allows 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 $e^{-i\omega t}$ and $e^{i\omega t}$. 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 (7) is done either by reformulating it as a quadratic program (Candès & Fernandez-Granda 2012a; Tang et al. 2013b; Chen & Chi 2013) or by discretizing the dictionary (Tang et al. 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 $\Delta\omega$, $\Omega = \{\omega_k = k\Delta\omega, k = 0..n\}$ and a $m \times 2n$ matrix A whose columns are $e^{-i\omega_k t}$ and $e^{i\omega_k t}$. In that case (9) reduces to:

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

Which is often referred to as the LASSO problem when W is the identity matrix. As the parameter λ is not so easy to tune, an equivalent formulation of discretized (9) is chosen,

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

where ε is a positive number. By "equivalent", we mean there exists a λ_ε such that the solution of (10) is equal to the solution of (11, BP$_{\varepsilon,W}$) (Rockafellar 1970). As this problem will often be referred to, we add to the equation number BP$_{\varepsilon,W}$ in the rest of the text, BP standing for Basis Pursuit. There are several codes written to solve (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 et al. 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]).

---

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

The solution of $(11, \mathrm{BP}_{\varepsilon,W})$ offers an estimate for the periods, but the efficiency of the method can be improved by using a moving average on $x^\star$, to approximate better (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 (5) $x^\star$ 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 $(11, \mathrm{BP}_{\varepsilon,W})$ selects significant frequencies in the data, but the estimates of their amplitude is biased due to the ℓ_1 norm minimization. To obtain unbiased amplitudes, we first check that the peaks are not aliases of each others. 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 W and the width η of the interval over which the result of $(11, \mathrm{BP}_{\varepsilon,W})$ is averaged.

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

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

(v) Denoting $\Omega = [\omega_{\min}, \omega_{\max}]$ for each frequency $\omega \in \{\omega_{\min} + \eta, ..., \omega_{\max} - \eta\}$, 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 $x^\sharp$.

(vi) Plot $x^\sharp$ as a function of the frequencies or the periods.

(vii) Evaluate the significance of the main peaks (figure 6).

Each of these steps are detailed in the following sections.

3.2 Optimization routine

Many solvers can handle (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 $\frac{|\|Ax-y\|_{\ell_2} - \varepsilon|}{\max\left(1, \|Ax-y\|_{\ell_2}\right)} <$ tol. The default parameters seem acceptable, in particular tol$=10^{-4}$.

3.3 Dictionary A

To estimate the spectrum, a natural choice for the columns of matrix A is $(\mathrm{e}^{-\mathrm{i}\omega t}, \mathrm{e}^{\mathrm{i}\omega t})$. 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 $((a_0 + a_1 t + a_2 t^2 + ...)\cos(\omega t + \phi))$ or Keplerian motions. In the latter

case, column vectors would be of the form $\frac{r}{a}\mathrm{e}^{\mathrm{i}\nu(t)}$ where $\nu(t)$ 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 (5) will be close to (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 $(11, \mathrm{BP}_{\varepsilon,W})$ 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 to mitigate the distortions induced by their removal.

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

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

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

3.5 Tuning

Choice of W: We have seen section 2.3 the weight matrix W is characterized by the covariance function R via $W_{kl} = R(|t_k - t_l|)$. Several forms for the covariance functions were suggested (e.g. Rajpaul et al. 2015). Here we only consider exponential covariances, that 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 noise are here supposed independent, the covariance function of their sum is the sum of their covariance functions. Therefore, the matrix 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/day as a default value but it is also advisable to re-do the analysis for 0.95 cycles/day, as discussed in the examples sections 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 $\|W(Ax - y)\|_{\ell_2} < \varepsilon$ that has the correct ℓ_0 norm. Let us consider a signal made of p pure sinusoids sampled at times $t = (t_k)_{k=1..m}$, $y(t) = \sum_{j=1}^{p} c_j \mathrm{e}^{\mathrm{i}\omega_j t}$. Assuming the frequencies on the grid are regularly spaced with step $\Delta\omega$, this leads

to the condition (see A for calculation details):

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

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, $\sqrt{\sum_{j=1}^{p}|c_j|^2} \approx \|y\|_{\ell_2}/\sqrt{m}$. 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) suggests 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 $\varepsilon_{\mathrm{grid}}$ and noise, which yields $\varepsilon_{\mathrm{noise}}$. Supposing the noise is Gaussian, denoting by y_t the underlying non noisy observations, $\|W(y_t-y)\|_{\ell_2}^2$ as a function of random variable $y = y_t + n$ follows a χ^2 distribution with m degrees of freedom. Denoting its cumulative distribution function (CDF) by $F_{\chi_m^2}$, the probability $1-\alpha$ that the true signal y_t is in the set $\{y', \|W(y'-y)\|_{\ell_2}^2 \leq \varepsilon_{\mathrm{noise}}\}$ is:

$$F_{\chi_m^2}(\varepsilon_{\mathrm{noise}}^2) = 1-\alpha \tag{16}$$

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

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

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 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 next section.

3.6 Post-processing

Once the solution to $(11, \mathrm{BP}_{\varepsilon,W})$ is computed, the spectrum $x^\star$ 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 $(11, \mathrm{BP}_{\varepsilon,W})$ might be untrustworthy. When the signal is made of several frequencies, the solution might over-estimate the one with the greatest amplitude, and under-estimate 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 $x^\star$ is the solution to $(11, \mathrm{BP}_{\varepsilon,W})$, denoting by $x^\star(\omega)$ the coefficient corresponding to frequency ω, we compute

$$\hat{y}_\omega(t) = \|Wy\|_{\ell_2}\sum_{\substack{\omega'\in\Omega \\ \omega-\eta\leqslant|\omega'|\leqslant\omega+\eta}}\frac{x^\star(\omega')a_{\omega'}(t)}{\|Wa_{\omega'}(t)\|_{\ell_2}} \tag{18}$$

Where $a_{\omega'}(t)$ is the column of A corresponding to frequency ω'. The terms $\|Wy\|_{\ell_2}$ and $1/\|Wa_{\omega'}(t)\|_{\ell_2}$ appear because the columns of WA and the weighted observations Wy were normalized in step 3.4. The vector $\hat{y}_\omega(t)$, $t = t_1..t_m$ 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 $x^\star$ 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, eq. 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, $\varepsilon^2 = F_{\chi^2}^{-1}(1-\tilde{\alpha})$. As $\varepsilon \approx \varepsilon_{\mathrm{noise}}$, 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 $(11, \mathrm{BP}_{\varepsilon,W})$ is not zero then a signal has been detected with a FAP lower 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 to reject the hypothesis that the signal is pure Gaussian noise of covariance matrix 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:

• Re-sampling: Taking off randomly 10-20% of the data and re-computing the ℓ_1-periodogram. The peaks that show great variability should not be trusted.

• 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

$$f_K(\theta_0, (\theta_{Kj})_{j=1..np}) = \text{non planetary}(\theta_0) + \sum_{j=1}^{n_p} \text{Keplerian}_j(\theta_{Kj}) \tag{20}$$

or

$$f_C(\theta_0, (\theta_{Cj})_{j=1..np}) = \text{non planetary}(\theta_0) + \sum_{j=1}^{n_p} \text{Circular}_j(\theta_{Cj}) \tag{21}$$

That is a sum of a model accounting for non planetary effects non planetary(θ_0), θ_0 being a real vector with n_θ components, and a sum of Keplerian or circular curves depending on five resp. three parameters, $\theta_{Kj} = (k_j, h_j, P_j, A_j, B_j)$ and $\theta_{Cj} = (P_j, A_j, B_j)$

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

$$\text{Circular}(\theta_C) = A\cos(\frac{2\pi t}{P}) + B\sin(\frac{2\pi t}{P}) \tag{23}$$

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-square 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:

$$\chi^2_{K,C}(\theta_0^{\text{fit}}, \theta_{n_p}^{\text{fit}}, \omega) =$$
$$\left[y - f_{K,C}\left(\theta_0^{\text{fit}}, \theta_{n_p}^{\text{fit}}, \omega^{\text{fit}}\right)\right]^T V^{-1} \left[y - f_{K,C}\left(\theta_0^{\text{fit}}, \theta_{n_p}^{\text{fit}}, \omega^{\text{fit}}\right)\right] \tag{24}$$

$$\chi^2_{K,C}(\theta_0^{\text{fit}}, \theta_{n_p}^{\text{fit}}) =$$
$$\left[y - f_{K,C}\left(\theta_0^{\text{fit}}, \theta_{n_p}^{\text{fit}}\right)\right]^T V^{-1} \left[y - f_{K,C}\left(\theta_0^{\text{fit}}, \theta_{n_p}^{\text{fit}}\right)\right] \tag{25}$$

$f_{K,C}\left(\theta_0^{\text{fit}}, \theta_{n_p}^{\text{fit}}, \omega^{\text{fit}}\right)$ is the model fitted depending on the non planetary effects θ_0, the (Keplerian or circular) $\theta_{n_p} = (\theta_{K,Cj})_{j=1..n_p}$ parameters of n_p planets plus a circular or Keplerian orbit initialized at frequency ω. V designates the covariance matrix of the noise model ($V^{-1} = W^2$ 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 equations (22),(23), only $(A_j)_{j=1..n_p+1}$ and $(B_j)_{j=1..n_p+1}$ are fitted and the non planetary effects are modelled linearly: there exists a matrix ϕ such that non planetary$(\theta_0) = \phi\theta_0$. 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:

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

Where ω_{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 is only non planetary effects and n_p planets. Baluev (2008) has computed tight bounds for that quantity in case of a circular model and a linear fit (corresponding to subscript C), which we reproduce here:

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

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 $W = \omega_{\max}\sqrt{(\bar{t^2}-\bar{t}^2)/\pi}$, t being the array of measurement times and $\bar{t}$ is the mean value of 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 periodogram defined by Zechmeister & Kürster (2009). In each case, the algorithm is tuned in the following way:

• The problem $(11, \mathrm{BP}_{\varepsilon,W})$ is solved with SPGL1 (van den Berg & Friedlander 2008)
• The solution of SPGL1 is averaged on an interval $\eta = 2\pi/(3T_{\mathrm{obs}})$ according to section 3.5.
• 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 section 3.7.4 and are represented figure 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 square 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 Generalized Lomb-Scargle 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 days. The precision of the measurements given in the raw data set (from now on called nominal precision) is between 0.8 and 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, 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 $(11, \mathrm{BP}_{\varepsilon,W})$ and average the solution as explained in 3.6.

The resulting array $x^\sharp(\omega)$ (see equation (19)) is plotted versus frequency, here giving figure 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 – figure 6.a. The FAP corresponding to a false alarm probability of 10^{-4} is represented by a dotted line.

The values of most of the algorithm parameters defined section 3.5 are fixed in the previous section. Here we precise that the method is performed for two grid spans: 0 to 1.5 cycles per day and 0 to 0.95 cycles/day (figure 1.b resp. c).

We first apply the method on a grid spanning between 0 and 1.5 cycles per day. The weight matrix is diagonal, $W_{kk} = 1/\sigma_k$ (not $1/\sigma_k^2$) where σ_k is the error on measurement k. On figure 1.b, the peaks of published planets appear, as opposed to the generalized Lomb-Scargle periodogram (1.a). However, there are still peaks around one day. 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 $\omega_M = 6.30084$ radian/day. 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 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 figure 6.a. Taking 8.667, 31.56 and 197 days gives a reduced χ^2 of the Keplerian fit with three planets plus a constant (16 parameters) is 1.19, yet the stellar jitter is not included. As a consequence, finding other significant signals is unlikely.

Looking only at figure 1.b, whether the signal at 197 days or its alias at 0.9921 days 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 days is 1.2548, suggesting the planet at 197 is in 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 per day to avoid spurious peaks (figure 1.c). As said above, the star is expected to have a jitter in the meter per second range, so we take for the additional jitter $\sigma_W = 0$, $\sigma_R = 1$ m/s and try several characteristic correlation time lengths $\tau = 0, 3, 6, 10$ or 20 days with definitions of equation (14). In that case, as said section 2.3, the estimation of the power is expected to be biased. Figure 1.c shows that the peaks at high and low frequencies are respectively over-estimated and under-estimated. We suggest the following explanation: the weighting matrix accounts for red noise that has more power at low frequencies. Therefore, the minimization of (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
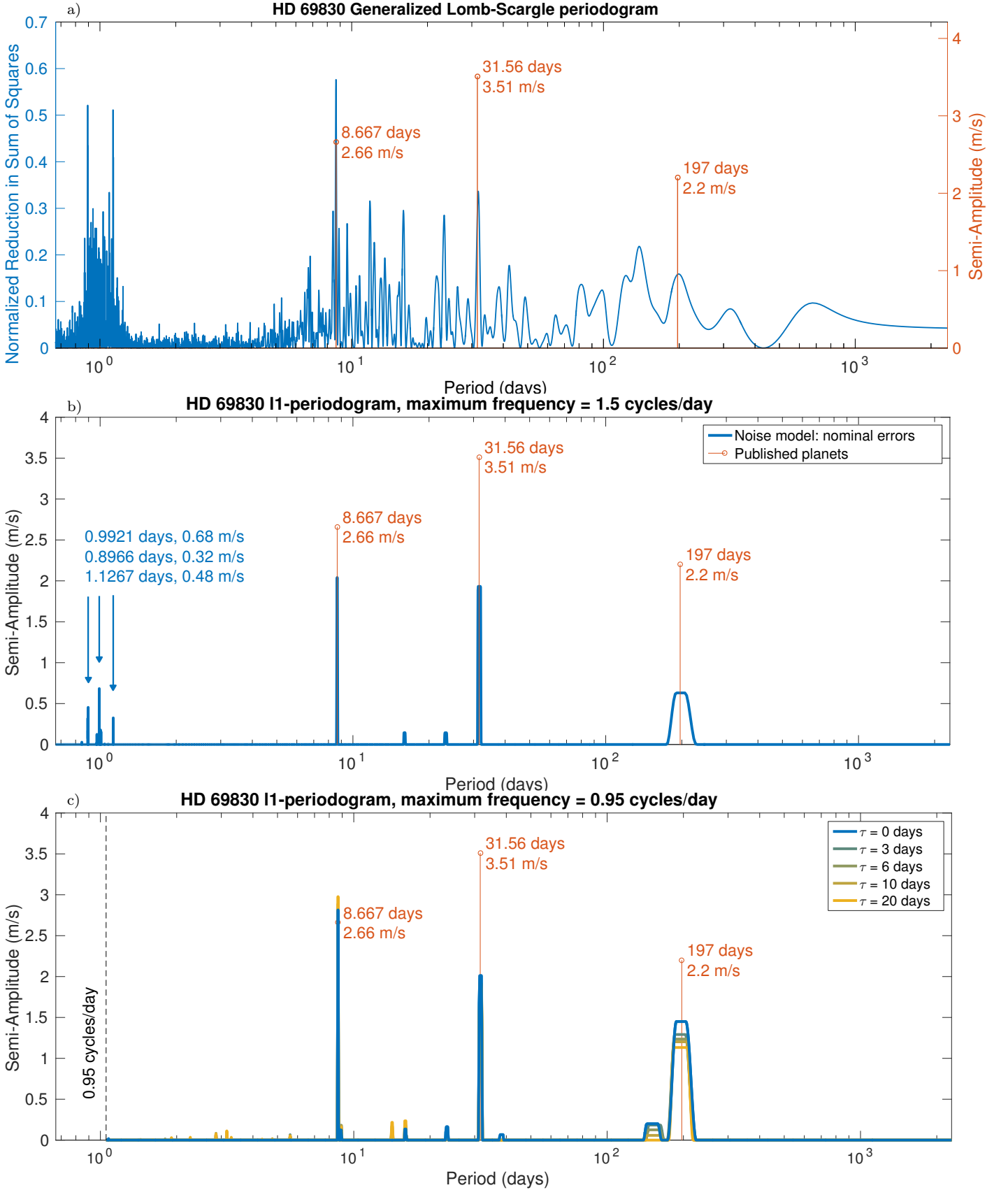
Figure 1. Generalized Lomb-Scargle periodogram and ℓ_1-periodogram of HD 69830 in blue, published planets are represented by the red stems. The frequency span used for figures b anc c are respectively 1.5 and 0.95 cycles/day. The other signals mentioned section 1 are spotted by the blue arrows. For all the noise model considered for matrix W, $\sigma_W = 0$, $\sigma_R = 1$ m.s^{-1}.
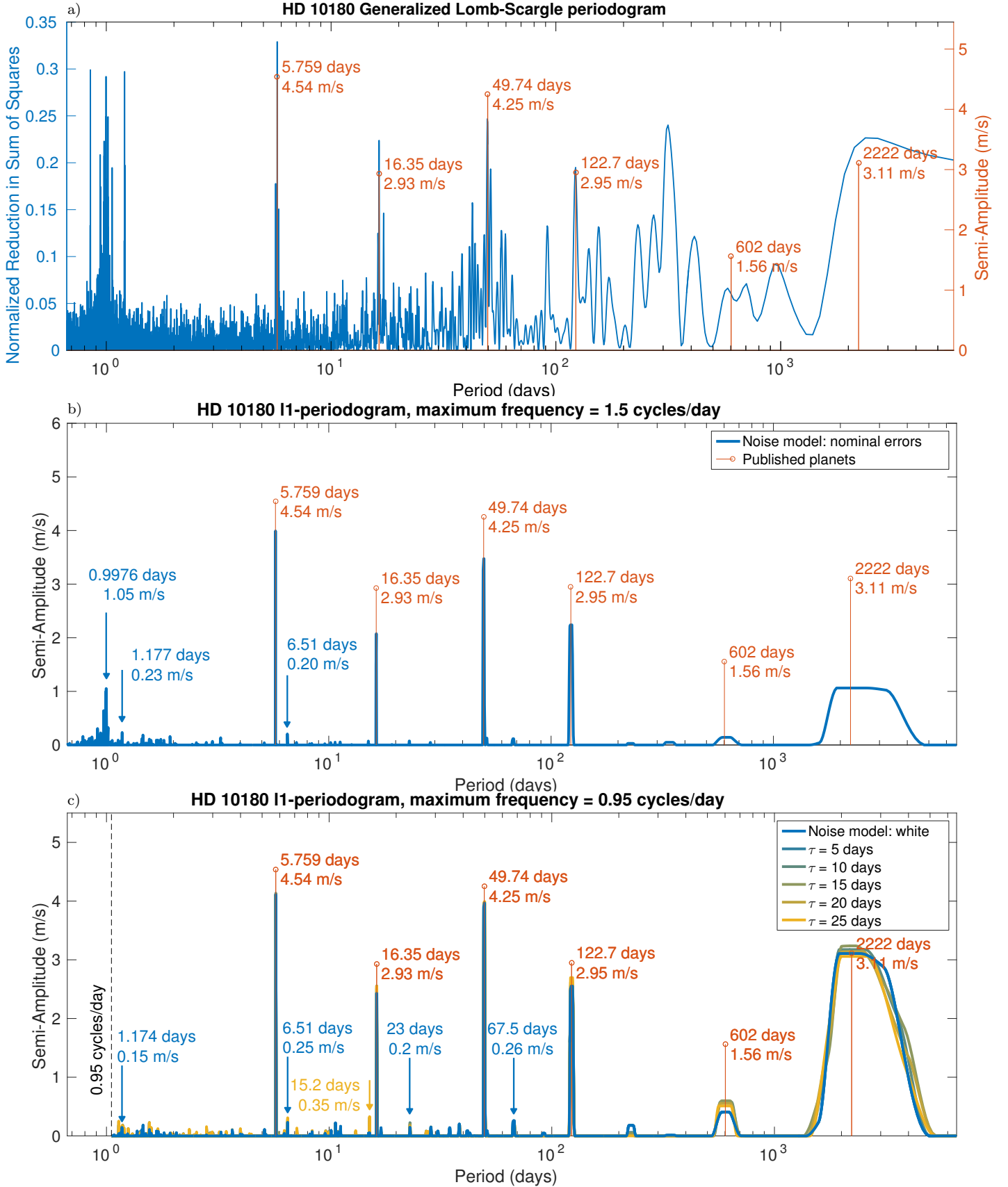
**Figure 2.** GLS and ℓ_1-periodograms of HD 10180 data set with mean subtracted. The red stems have the periods and amplitude of published planets. The other signals mentioned section 2 are spotted by the blue arrows. For all the noise model considered for matrix W, $\sigma_W = 0$, $\sigma_R = 1$ m.s^{-1}.
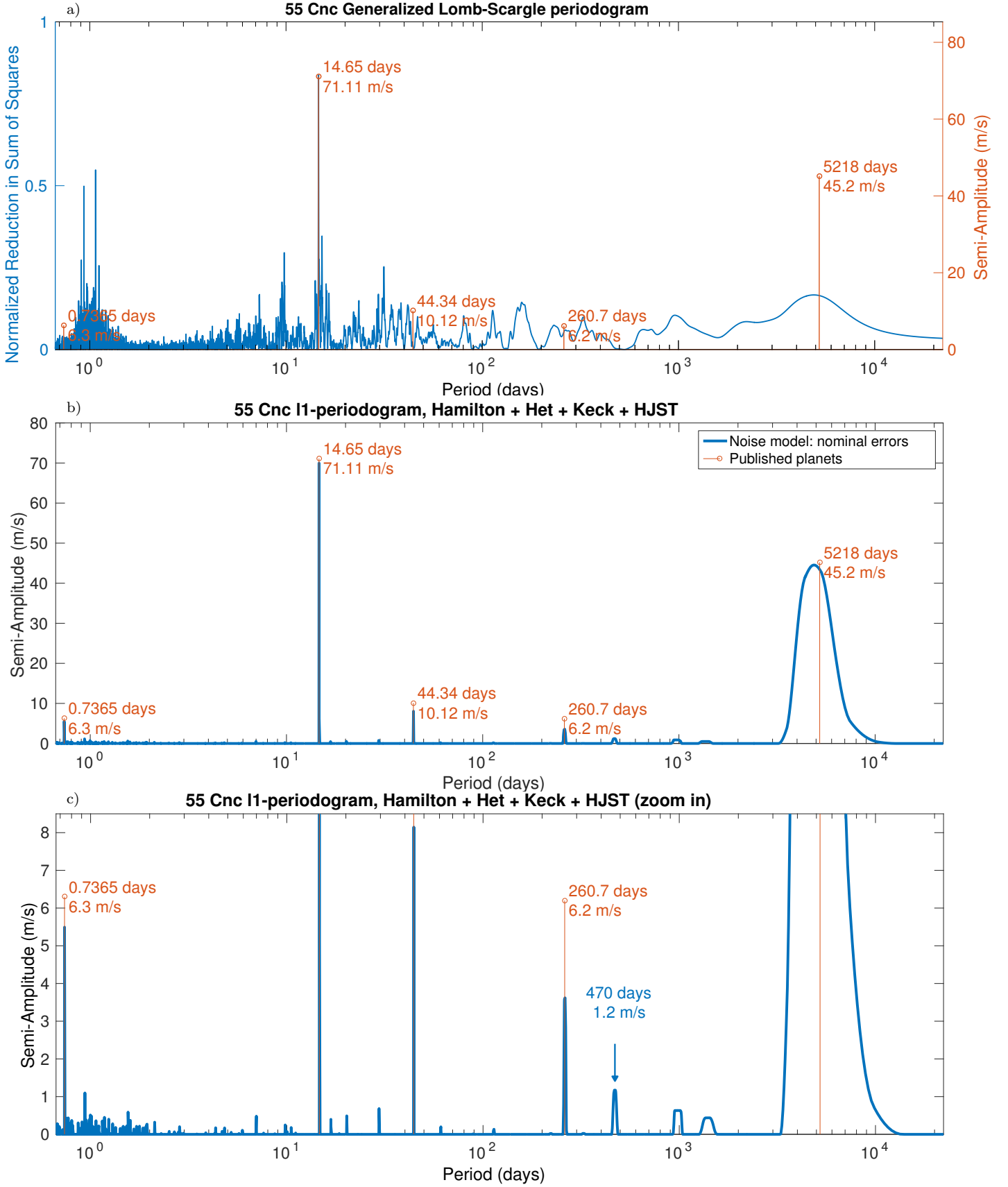
Figure 3. GLS and ℓ_1-periodograms of 55 Cnc data set with mean subtracted. The red stems have the periods and amplitude of published planets. The other signals mentioned section 4.4 are indicated by the blue arrows.
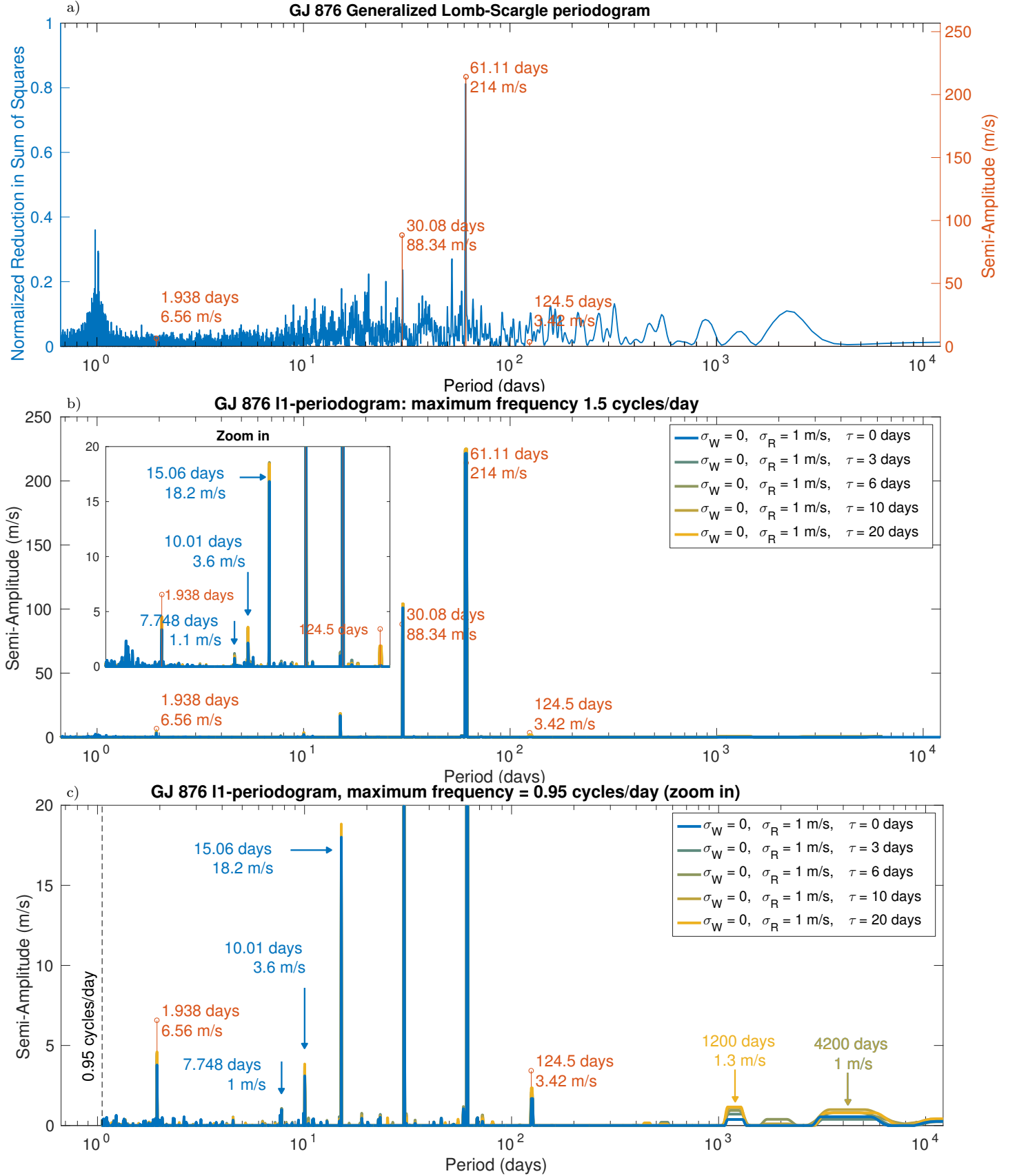
Figure 4. GLS and ℓ_1-periodograms of GJ 876 data set with means of KECK and HARPS measurement respectively subtracted. The red stems have the periods and amplitude of published planets. The other signals mentioned section 4.5 are indicated by the blue arrows.
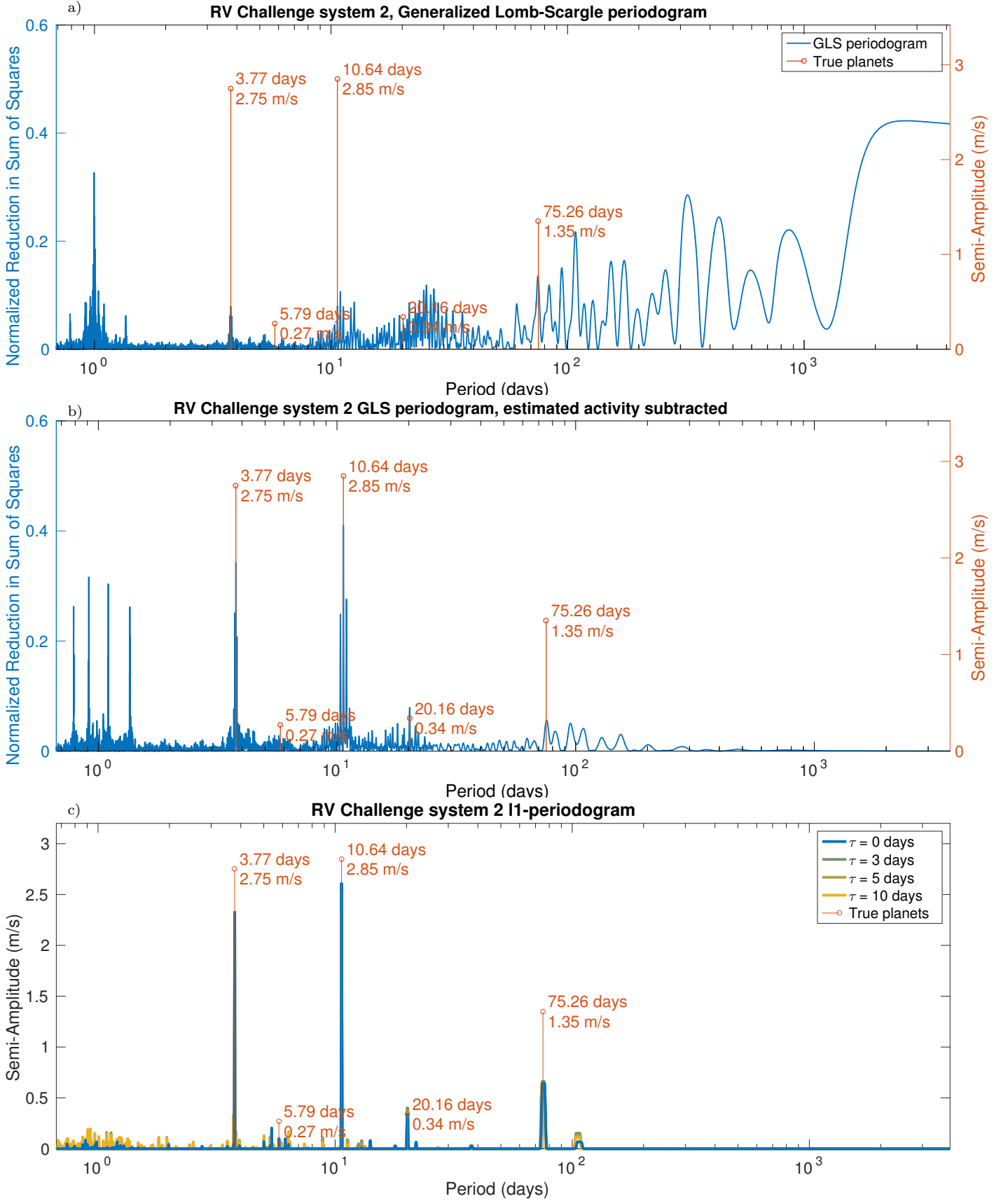
Figure 5. Top: GLS of the RV Challenge system 1 (simulated signal). Top: GLS of raw data, middle: GLS after fitting ancillary measurements, bottom: ℓ_1-periodogram after fitting ancillary measurements. True planets are represented by red lines.
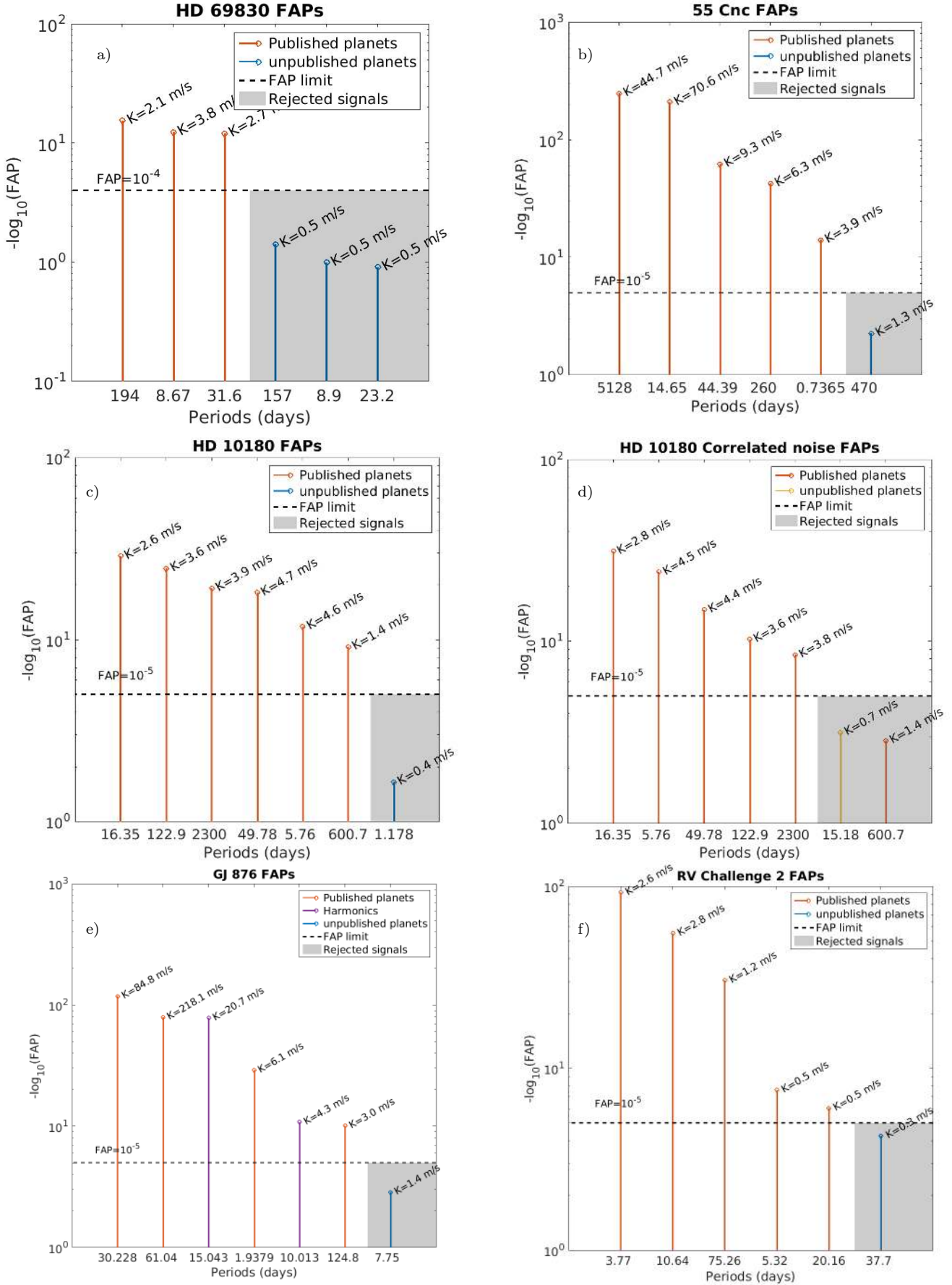
Figure 6. Peak amplitudes and associated FAPs for the four systems analysed   MNRAS **000**, 1–28 (2016)

periodogram does not correspond to a planet in $\approx 7\%$ of the cases, while the maximum peak of the ℓ_1-periodogram is spurious in less than 0.5% 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 is such that $\log R'_{HK} = -5$ 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 days is firmly stated. Let us mention that there is a concern on whether a planet at 227 days could be in the signal instead of 600 days, as they both appear on the periodogram of the residuals and $1/227 - 1/600 + 1/365 \leqslant 1/T_{\text{obs}}$, where T_{obs} is the total observation time. The possibility of the presence of a seventh planet 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 days (Lovis et al. 2011). They are such that $1/6.51 + 1/1.178 - 1 \leqslant 1/T_{\text{obs}}$, so one is probably the alias of the other. The dynamical stability of a planet at 1.17 days is discussed in Laskar et al. (2012), and its ability to survive is shown. However in our analysis, the statistical significance is too low to claim the planet is actually in the system.

We compute the ℓ_1-periodogram for a grid span of 0 to 1.5 cycle/day and 0 to 0.95 cycles per day, giving respectively figures 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 appendix B and previous section, in that case we have to decrease $\varepsilon_{\text{noise}}$ and here $F_{\chi^2_m}(\varepsilon^2_{\text{noise}}) = 0.1$ was taken. Where $F_{\chi^2_m}$ is the cumulative distribution function of the χ^2 distribution with m degrees of freedoms, 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 days and that there is a small peak at 23 days, which is close to the stellar rotation period estimate of 24 days (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 per day. In that case the highest one at 0.9976 days corresponds to an alias of the 2222 days period. We will see that in the two next systems the aliases are not as disturbing, which is discussed section 5.2.

We now need to evaluate the significance of the peaks. The FAP test is performed for the seven highest signals, that are the published planets plus 0.177 days or 15.2 days. The latter appears for a non-diagonal weight matrix 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 $\tau = 25$ days (with notations of equation (14)). This analysis gives figure 6.c and d. In both cases the signals are below the significance threshold. It is also not clear which seventh signal to choose (figure 2.c), but doing the analysis with other candidates as 6.51, 23 or 67.5 days does not spot significant signals either. Let us note that when choosing a non diagonal W, the FAP of the 16.4 and 600 days planets respectively increase and decrease. We sug-

gest 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$_j$ minimum mass planet at 14.7 days 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 days and a possible Jupiter mass companion at 44.3 days. With the same obsevations and additional ones from the Hobby-Eberly Telescope (HET) and ELODIE, McArthur et al. (2004) suggested a Neptune mass planet could be responsible for a 2.8 days period. Wisdom (2005) re-analysed the same data set and found evidence for a Neptune-size planet at 261 days and suggested that the 2.8 period is spurious. This was confirmed by Dawson & Fabrycky (2010), which showed that the 2.8 days periodicity is an alias and the signal indeed comes from a super-Earth orbiting at 0.7365 days. 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 days 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 days 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 (Eberly Telescope), giving planets at 0.736546 ± 3.10^{-6}, 14.651 ± 10^{-4}, 44.38 ± 7.10^{-3}, 261.2 ± 0.4 and 4909 ± 30 days. 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 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. Figure 3.b shows the ℓ_1-periodogram and 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 comes 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 days periodicity clearly appears on the ℓ_1-periodogram.

The FAPs computed following the method outlined sec-
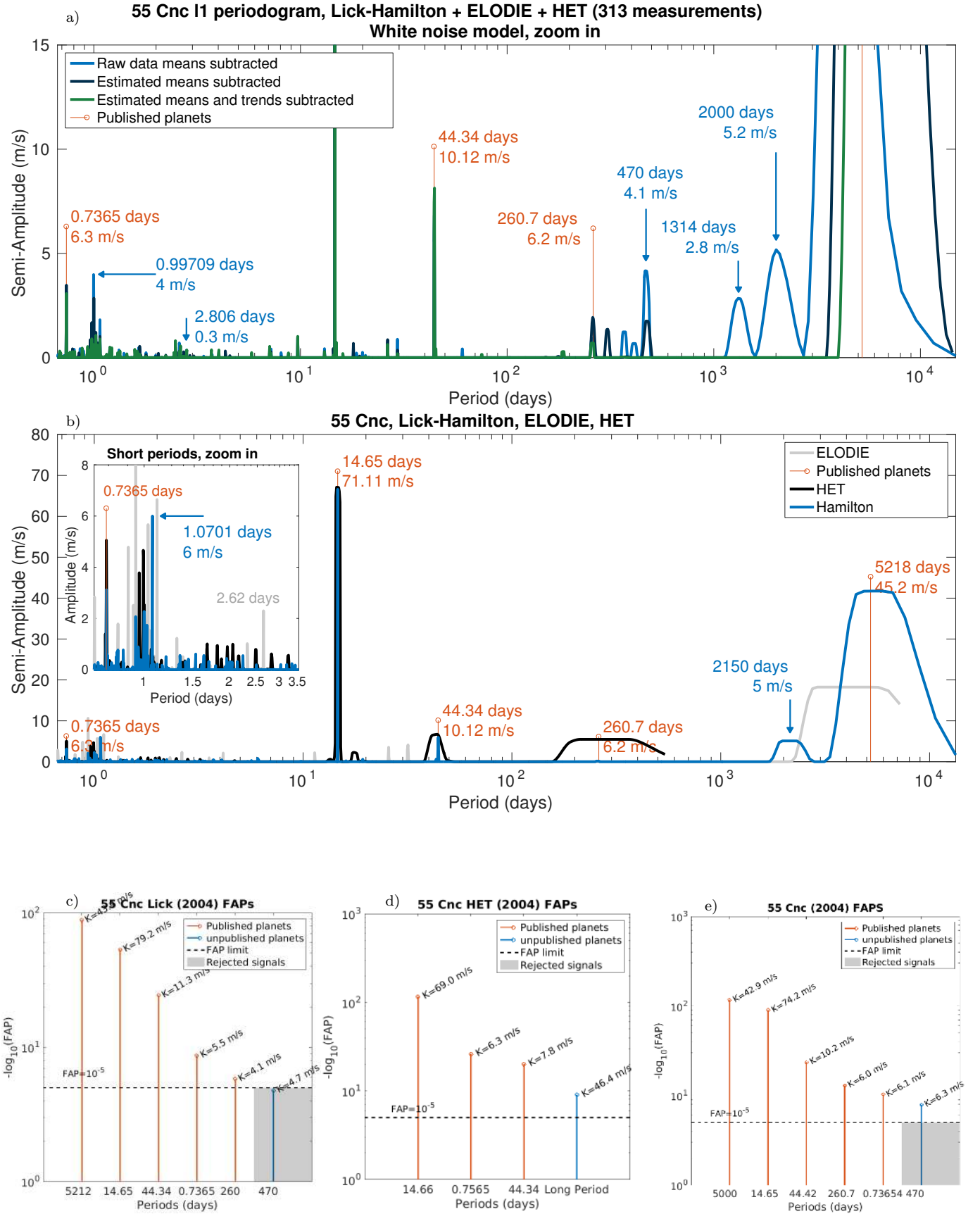
Figure 7. ℓ_1-periodogram of 55 Cnc, using measurements from the Lick-Hamilton, ELODIE spectrograph (Observatoire de Haute provence) and HET telescope.
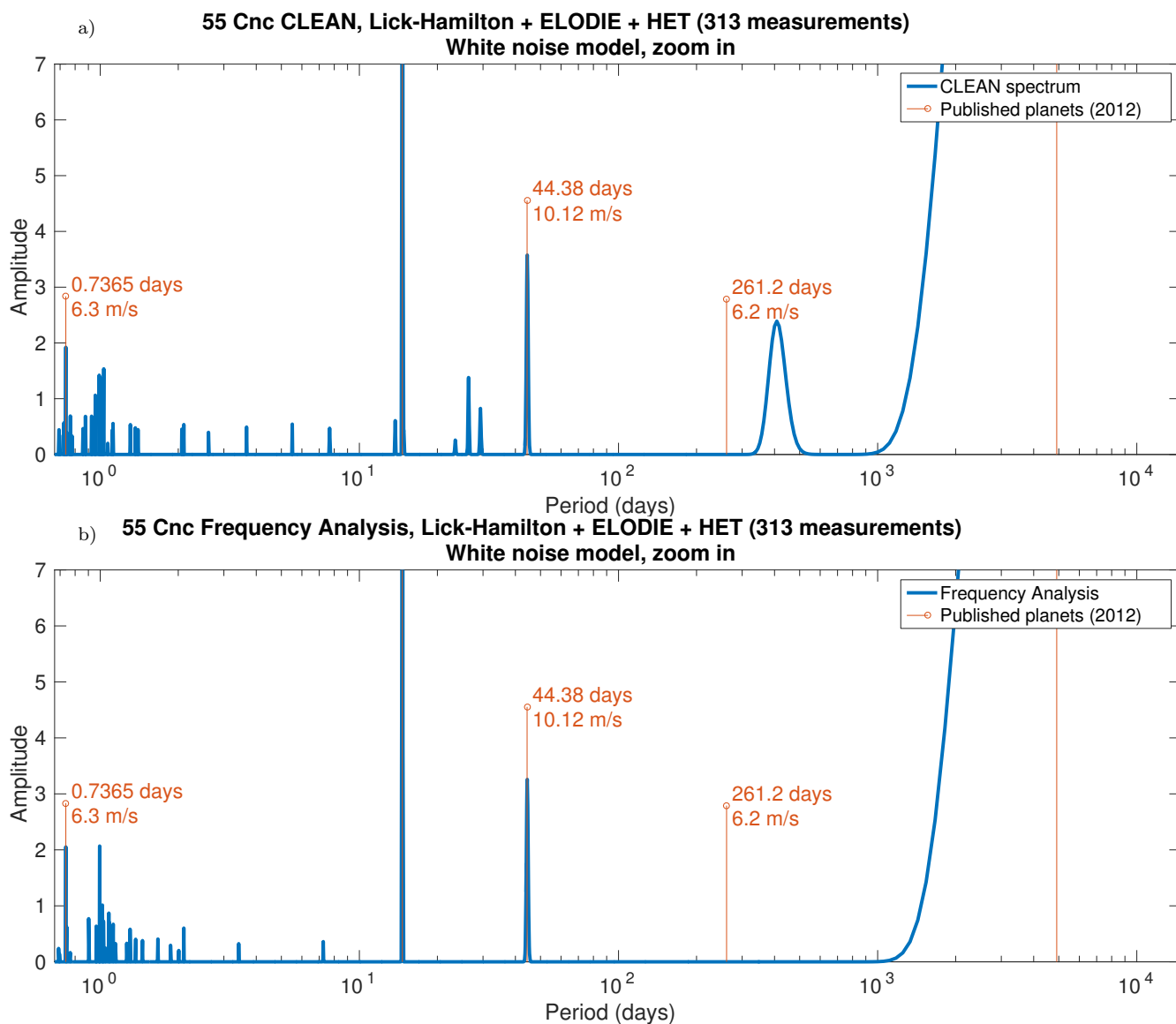
Figure 8. a) CLEAN spectrum of 55 Cnc with the data available in 2004, b) Frequency Analysis of the same data

tion 3.7.4 are significant (see figure 6.b). The sixth highest peak is at at 470 days, 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 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 days in early measurements is also discussed.

4.4.2   Measurements before 2004: no planet at 2.8 days nor 470 days 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 days, and among candidate signals, one was confirmed (260 days) and one was not (470 days). 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 shorten 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 Hobby-Eberly Telescope (HET) and ELODIE (Observatoire de Haute Provence). We also show that the signal at 0.7365 days (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 on figure 7.a. The true periods appear, although the 260 period is very small and there are peaks at 470, 1314 and 2000 days (the other features of the figure will be explained later). We then consider the three data sets separately, the figure obtained is displayed figure 7.b. The fact that the ℓ_1-periodograms of each three instruments

span on different length is due to the fact that they don't 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 sets, but the 0.7365 and 44.34 days periodicities are not seen for the ELODIE data set. Interestingly, HET ℓ_1-periodogram displays a periodicity close to 260 days. However, one cannot claim a detection at this period in HET data, as those only span on 180 days, any period longer than the observation timespan is very poorly constrained. Furthermore, the period at 2.8 days is not seen in any data set. The closest one would be a peak at 2.62 days obtained with ELODIE data, which was checked not to be significant. The 470 days periodicity does not appear either. We show 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 figure 8) also allow 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 comes 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 days, but this one can be discarded as it is an alias of the 14.65 days periodicity. In both HET and Hamilton data, the 0.7365 periodicity is significant (figure 7.c and d). Also, one sees a significant long period in both cases (respectively 8617 and 5212 days). The HET data set spans on 170 days, 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 days periodicities become insignificant.

The difference in zero points of the three instruments has a signature on the ℓ_1-periodogram. Indeed, in problem $(11, {\rm BP}_{\varepsilon,W})$, 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 ≈ 14.65 day, 44.34, 5000 and 0.7365 days. The fit with these four periods plus the vectors $1_I(t)$ gives coefficients of the latter $\alpha_{\rm Lick}, \alpha_{\rm Elodie}$ and $\alpha_{\rm HET}$. The vector $\alpha_{\rm Lick}1_{\rm Lick}(t) + \alpha_{\rm Elodie}1_{\rm Elodie}(t) + \alpha_{\rm HET}1_{\rm HET}(t)$ is subtracted from the raw data. The ℓ_1-periodogram of the residuals is computed, which gives the dark blue curve figure 7.a). The 2000 and 1314 periods disappear and the 470 days peaks decreases. Interestingly enough the 5th tallest peak (except the 0.99709 days alias) becomes 260 days, 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 (figure 8.a and b).

We now fit the model with five planets along with the 1_I vectors and trends for each instrument, that are vectors t_I such that $t_I(t) = t$ and 0 elsewhere if the measurement at time t is done by the instrument I. The vector $\sum \alpha_k 1_{I_k} + \beta_k t_{I_k}$ is subtracted from the raw data, and we compute again the

ℓ_1-periodogram (figure 7.a, green curve). This time, the 470 days periodicity disappears, suggesting – though not proving – it is due to a difference in behaviour between the instruments. The fact that the 470 days 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 days, 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 days 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 days, not ≈ 2.8 days as it was then. We found that the size of the peak in the residuals at 260 days 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, 15000 ($1/14.65 - 1/14.8 \approx 1/5000 - 1/15000$) to those of the four planets and a 2500 one (probably due to an harmonic of the 5000 days periodicity) makes the 400 (seen on the CLEAN spectrum figure 8.a) and 470 periodicity disappear and the 260 days peak appears very clearly. As the data comes 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 figure 7.e). The model is made of Keplerians plus the 1_I vectors. The four significant signals in each data set are still significant. The 260 days periodicity is significant as well. This analysis shows that both the 0.7365 and 260 days periodicity 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 contrary, it seems hard to explain a steady 0.7365 days periodicity with a non-planetary effect.

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 days were reported by (Marcy et al. 1998; Delfosse et al. 1998). Subsequently, Rivera et al. (2005) finds a short period Neptune at 1.94 days and a Uranus-mass planet at 124 days (Rivera et al. 2010).

The giant planets are close to each others 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 4-body Newtonian fits which give a χ^2 of the residuals smaller than a Keplerian fit. The dynamical fits also allow 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 $\tau = 3$ days.

Jenkins et al. (2014) takes a different approach and searches for sine functions in the signal. They claim 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 they are likely to be due to the interactions between the giant planets.

4.5.2 Six significant sines

Jenkins et al. (2014) analyses the GJ 876 data by aiming at solving the problem (1), which they call Minimum Mean Squared Error (MMSE). 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 days. 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 cycles per day and 0.95 cycles/day (see figure 4 b and c). On figure 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 days as candidate 7th planets, respectively with the W matrix yielding their greatest amplitude. On figure 6.e), we display the result for 7.748 days but in other cases the signals are not significant. Let us still point out that in the case of $\tau = 6$ days, initializing a 4200 days periodicity, after the non-linear fit we obtain a 4862 days periodicity which has a FAP of 0.0007. This one is close to the total observation timespan (4600 days). Therefore it is hard to determine what could be its cause.

Before discussing the origin of these signals, we wish to comment the behaviour of the ℓ_1-periodogram towards the 124 days perodicity. Indeed, in the case of the 1.5 cycles per day, this one has the same order of magnitude as the tallest alias in the one day region (at 0.9812 days, alias of the 61 days periodicity). Furthermore, the peak becomes visible only for non diagonal weight matrix W, while a white noise model is sufficient to see it when using a shorter grid (figure 4.c). 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 one day 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 section 4.4.2. In that case the 124 days periodicity does appear and the aliases are reduced. We suggest the following explanation: when at least one of the three obstacle 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 days

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

We perform the same 4-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 days the inclination is also a free parameter.

A planetary system with the orbital elements found by the least square fit is simulated on 100 years for the two giant planets and the four planets at once. The frequency analysis (Laskar 1988; Laskar et al. 1992; Laskar 1993) 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 planets 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 days 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 days could be due to additional planets, and find it unlikely. They compute the evidence ratio of Newtonian models with four and five planets, $\mathrm{Pr}\{y|5\text{ planets}\} / \mathrm{Pr}\{y|4\text{ planets}\}$, and find 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 $2\pi(1/\omega_{15} - 2/\omega_{30}) \approx 8.77$ years, which is consistent with the estimate of Correia et al. (2010) ($g_2 = 8.73$ year, 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 period 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 has not a 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 et al. (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) insofar as we use the ancillary measurements as proxys for estimating the activity induced signal. Here, we simply fit and remove the three ancillary measurements from the data then use the method described above on the residuals. To compute the FAP we use a model of the form $A \mathrm{FWHM} + \mathrm{B} \mathrm{bisector} + \mathrm{C} \log \mathrm{R}'_{\mathrm{HK}} + \mathrm{Circ}(\mathrm{k}, \mathrm{h}, \mathrm{P}, \mathrm{D}, \mathrm{E})$, Circ denoting a circular model as defined section 3.7.3. The validity of this approach is discussed in D.

The data set used is taken from the RV Fitting Challenge (Dumusque 2016; Dumusque et al. 2016). In this challenge, fifteen systems were simulated with a red noise component taken from observations of real stars plus activity simulated via SOAP 2 (Dumusque et al. 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 (6 parameters). Secondly, we compute the ℓ_1-periodogram for different weight matrices, which gives figure 5.c. The Generalized Lomb-Scargle periodogram is also computed before and after the fit of the 6 parameters for comparison (figure 5.a and b).

We find without ambiguity the three planets whose semi-amplitude is above 1 m.s^{-1}, and also the 20.16 days periodicity. The planet with the smallest amplitude does not appear clearly, but there is a peak at 5.4 days which seems to be significant. In fact, the spectral window is such that 5.4 days is an alias of $5.32 = 10.64/2$ days, 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 amplitude of 5.32 and 5.4 changes in favour of 5.32 days. 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 days 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 W. One of the main advantages of the method is that, as opposed to global χ^2 minimization, the minimization problem $(11, \mathrm{BP}_{\varepsilon, W})$ is convex therefore quicker to solve. On our workstation (Intel Xeon CPU E5-2698 v3 at 2.30 GHz) it takes typically thirty seconds to ten minutes 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 A (see section 3.3), n, and the precision wanted in output, tol (see section 3.2). The SPGL1 algorithm used to solve $(11, \mathrm{BP}_{\varepsilon, W})$ 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 onto 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 square minimisation. 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 Generalized Lomb-Scargle 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 is there a way to determine which peaks are to be associated to 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 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 one day 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 follow. Considering a set of observational times $t = t_1...t_m$, a linear combination of p pure sine signals $y(t_k) = a_1 \cos(\omega_1 t_k + \phi_1) + ... + a_p \cos(\omega_p t_k + \phi_p)$ is generated with uniformly distributed phases ϕ and various amplitudes. For any tolerance ε, the SPGL1 algorithm must give a solution $x^\star$ (see equation $(11, \mathrm{BP}_{\varepsilon, W})$) such that $\|x^\star\|_{\ell_1} \leqslant |a_1| + ... + |a_p|$, as obviously $y(t)$ belongs to the set of signals u verifying $\|u - y(t)\|_{\ell_2} \leqslant \varepsilon$. To test if SPGL1 gives the best solution we take the measurement dates of HD 69830 and generate three pure cosine functions of amplitude one whose frequencies are in the grid. They are fed to the SPGL1 solver for $\varepsilon = 0.01$ and W equal to the identity matrix. The solution $x^\star$ to $(11, \mathrm{BP}_{\varepsilon, W})$ must verify $\|x^\star\|_{\ell_1} < 3$ as the original signal is not noisy. The test is performed for a thousand set

Table 1. Maximum amplitude of the spectral window in the 1 cycle/day and 1 cycle per year for the examples considered here

	≈ 1 cycle/day	≈ 1 cycle/year
HD 69830	0.926	0.600
HD 10180	0.949	0.703
55 Cnc	0.822	0.557
GJ 876	0.73246	0.501
RV Challenge 2	0.870	0.800

of three frequencies randomly selected on the grid. We find that the average ℓ_1 norm of the solution is 3.26, suggesting 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.

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

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

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

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 one day, 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 figures 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 multi-tone 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 minimisation (1) for any number of sines inferior to the number of measurements and regarding their amplitudes, which ones seem to truly be 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 (Anglada-Escudé & Tuomi 2012), the treillis approach (Jenkins et al. 2014) or the super-resolution methods (Candès & Fernandez-Granda 2012b; Tang et al. 2013b) can be viewed as a way to approximate (1) and selecting the relevant number of frequencies at the same time. These ones 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 is 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 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 are more reliable in general than orthogonal matching pursuit. A failure of the ℓ_1-periodogram is also informative, as shown figure C2 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 short-cut to random searches and its use decreases the chance of being mislead 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 (9); for example by discretizing the dictionary, and ensuring the grid spacing is consistent with the noise power (see eq (15)) then solving (11, $\text{BP}_{\varepsilon,W}$) with SPGL1 and take the average power. The introduction of the weight matrix 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 systems 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 short-cut 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 are two-fold. 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 W must be studied into more depth. Secondly, let us recall that our method uses an *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:

• Adapting the method for very eccentric orbits, through the addition of Keplerian vectors to the dictionary for example.

• 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.

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

7 ACKNOWLEDGEMENTS

REFERENCES

Aigrain S., Gibson N., Roberts S., Evans T., McQuillan A., Reece S., Osborne M., 2011, in AAS/Division for Extreme Solar Systems Abstracts. 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 Processing, 98, 275
Babu P., Stoica P., 2010, Digital Signal Processing, 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, Celestial Mechanics and Dynamical Astronomy, 111, 235
Baluev R. V., 2013a, Astronomy and Computing, 3, 50
Baluev R. V., 2013b, Monthly Notices of the Royal Astronomical Society, 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 Journal on Imaging Sciences, 4, 1
Bellmann K., 1975, Biometrische Zeitschrift, 17, 271
Bobin J., Starck J.-L., Ottensamer R., 2008, IEEE Journal of Selected Topics in Signal Processing, 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., 2012a, preprint, (arXiv:1211.0290)
Candès E. J., Fernandez-Granda C., 2012b, CoRR, abs/1203.5871
Candès E., Romberg J., Tao T., 2006a, Information Theory, IEEE Transactions on, 52, 489
Candès E. J., Romberg J. K., Tao T., 2006b, Communications on Pure and Applied Mathematics, 59, 1207
Chandrasekaran V., Recht B., Parrilo P. A., Willsky A. S., 2010, preprint, (arXiv:1012.0621)
Chen Y., Chi Y., 2013, preprint, (arXiv:1304.4610)
Chen S. S., Donoho D. L., Saunders M. A., 1998, SIAM JOURNAL ON SCIENTIFIC COMPUTING, 20, 33
Cohen A., Dahmen W., Devore R., 2009, J. Amer. Math. Soc, pp 211–231
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., GÃijntÃijrk C. S., 2010, Communications on Pure and Applied Mathematics, 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, Information Theory, IEEE Transactions on, 52, 1289
Donoho D. L., Elad M., Temlyakov V. N., 2006, IEEE TRANS. INFORM. THEORY, 52, 6
Duarte M. F., Baraniuk R. G., 2013, Applied and Computational Harmonic Analysis, 35, 111
Dumusque X., 2016, The Radial Velocity Fitting Challenge. I. Simulating the data set including realistic stellar radial-velocity signals, Submitted to A&A
Dumusque X., Boisse I., Santos N. C., 2014, ApJ, 796, 132
Dumusque X., et al., 2016, The Radial Velocity Fitting Challenge. II. First results of the analysis of the data set, Submitted to A&A
Endl M., et al., 2012, ApJ, 759, 19
Engelbrecht C. A., 2013, in Precision Asteroseismology. pp 77–84, doi:10.1017/S1743921313014129, http://journals.cambridge.org/article_S1743921313014129
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 Processing, pp 600–616
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 Limited, pp 95–110
Gregory P. C., 2011, MNRAS, 410, 94
Gregory P. C., 2016, preprint, (arXiv:1601.08105)
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, Inc.

Kay S., Marple S.L. J., 1981, Proceedings of the IEEE, 69, 1380

Laskar J., 1988, A&A, 198, 341

Laskar J., 1993, Celestial Mechanics and Dynamical Astronomy, 56, 191

Laskar J., 2003, in proceedings of Porquerolles School, sept. 2011.

Laskar J., Froeschlé C., Celletti A., 1992, Physica D Nonlinear Phenomena, 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

Mallat S. G., Zhang Z., 1993, IEE Transactions on Signal Processing

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

McArthur B. E., et al., 2004, ApJ, 614, L81

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 Electrical and Electronics Engineers in Israel, 2008. IEEEI 2008. IEEE 25th Convention of. pp 290–294, doi:10.1109/EEEI.2008.4736707

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 Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on. pp 40–44 vol.1, doi:10.1109/ACSSC.1993.342465

Pelat D., 2013, Bases et méthodes pour le traitement de données

Queloz D., et al., 2001, A&A, 379, 279

Rajpaul V., Aigrain S., Osborne M. A., Reece S., Roberts S. J., 2015, preprint, (arXiv:1506.07304)

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, The Astrophysical Journal, 719, 890

Roberts D. H., Lehar J., Dreher J. W., 1987, AJ, 93, 968

Rockafellar R. T., 1970, Convex Analysis

Scargle J. D., 1982, ApJ, 263, 835

Schuster A., 1898, Terrestrial Magnetism, 3, 13

Schwarzenberg-Czerny A., 1998, Baltic Astronomy, 7, 43

Ségransan D., et al., 2011, A&A, 535, A54

Starck J.-L., Elad M., Donoho D. L., 2005, IEEE Transactions on Image Processing, 14, 1570

Stoica P., Babu P., 2012, Signal Processing, 92, 1580

Sulis S., Mary D., Bigot L., 2016, preprint, (arXiv:1601.07375)

Tang G., Bhaskar B., Recht B., 2013a, in Signals, Systems and Computers, 2013 Asilomar Conference on. pp 1043–1047, doi:10.1109/ACSSC.2013.6810450

Tang G., Bhaskar B., Shah P., Recht B., 2013b, Information Theory, IEEE Transactions on, 59, 7465

Tibshirani R., 1994, Journal of the Royal Statistical Society, Series B, 58, 267

Tropp J. A., Gilbert A. C., 2007, IEEE TRANS. INFORM. THEORY, 53, 4655

Tropp J. A., Laska J. N., Duarte M. F., Romberg J. K., Baraniuk R. G., 2009, CoRR, abs/0902.0026

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

Winn J. N., et al., 2011, ApJ, 737, L18

Wisdom J., 2005, in AAS/Division of Dynamical Astronomy Meeting #36. p. 525, http://web.mit.edu/wisdom/www/planet.pdf

Zechmeister M., Kürster M., 2009, A&A, 496, 577

Zucker S., 2015, MNRAS, 449, 2723

Zucker S., 2016, preprint, (arXiv:1601.01225)

van den Berg E., Friedlander M. P., 2008, SIAM Journal on Scientific Computing, 31, 890

APPENDIX A: MINIMUM GRID SPACING

Let us consider a signal made of p pure harmonics sampled at times $t = (t_k)_{k=1..m}$, $y = \sum_{j=1}^{p} c_j e^{i\omega_j t}$. We denote by ω_j' and $\Delta\omega$ two a 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,

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

$$= 2|c_j| \left| \sin\left(\frac{\omega_j - \omega_j'}{2} t_k \right) \right|$$

So denoting $y' = \sum_{j=1}^{p} c_j e^{i\omega_j' t}$,

$$|y_k - y_k'| = \left| \sum_{j=1}^{p} c_j \left(e^{i\omega_j t_k} - e^{i\omega_j' t_k} \right) \right|$$

$$\leqslant 2 \sum_{j=1}^{p} \left| c_j \sin\left(\frac{\omega_j - \omega_j'}{2} t_k \right) \right|$$

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

$$\|W(y - y')\|_{\ell_2}^2 \leqslant \|W\|^2 \|y - y'\|_{\ell_2}^2$$

$$\leqslant \|W\|^2 \sum_{k=1}^{m} |y_k - y_k'|^2$$

$$\leqslant 4\|W\|^2 \sum_{k=1}^{m} \left(\sum_{j=1}^{p} \left| c_j \sin\left(\frac{\omega_j - \omega_j'}{2} t_k \right) \right| \right)^2$$

given ((A3)),

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

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

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

given ((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 $\varepsilon_{\text{grid}}$ can be chosen as:

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

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

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

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 $(11, \text{BP}_{\varepsilon,W})$, reproduced here:

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

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 follow:

• The measurement times are those of HD 69830 (74 measurements);
• The true signal is $y(t) = 1\cos(\frac{2\pi}{7.5}t) + 2\cos(\frac{2\pi}{40}t + 2) + 2\cos(\frac{2\pi}{120}t + 1)$ m.s^{-1}.
• The noise is red, with parameters $\sigma_W = 0$, $\sigma_R = 2$ m.s^{-1} and $\tau = 12$ days, where σ_W, σ_R and τ are the parameters of the autocorrelation function R defined equation ((14)) reproduced here:

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

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, fifty noise time series $(n_k(t))_{k=1..50}$ are generated and the method

is applied to each $y_k(t) = y(t) + n_k(t)$ for three different weight matrices, all other parameters being fixed. In each case they are defined according to model (14) with $\sigma_W = 0$, $\sigma_R = 2$ m.s^{-1} and $\tau = 0$, 6 or 12 days. The grid goes between 0 and 0.95 cycles/day and ε verifies $F_{\chi_m^2}(\varepsilon_{\text{noise}}^2) = 0.1$. The resulting ℓ_1-periodograms are averaged (see figure B1.b).

To compare with a classical approach, we also compute classical periodograms for the same signals $y_k(t)$ 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 $y(t)$ with the same weight matrices as the ones used above. This gives figure 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 try of different noise matrices on real radial velocity data sets to see if a peak appears. Secondly, when 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 days periodicity is then under-estimated. This bias could prevent 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 under-estimated decreases. This is why we took $\varepsilon_{\text{noise}}$ such that $F_{\chi_m^2}(\varepsilon_{\text{noise}}^2) = 0.1$ and not $F_{\chi_m^2}(\varepsilon_{\text{noise}}^2) = 0.999$, 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 to chose an appropriate weight matrix automatically. In all cases when the algorithm has converged we have $\|W(Ax - y)\|_{\ell_2} = \varepsilon$ 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 WA and the weighted observations Wy are normalized. Therefore the problem always comes down to minimizing

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

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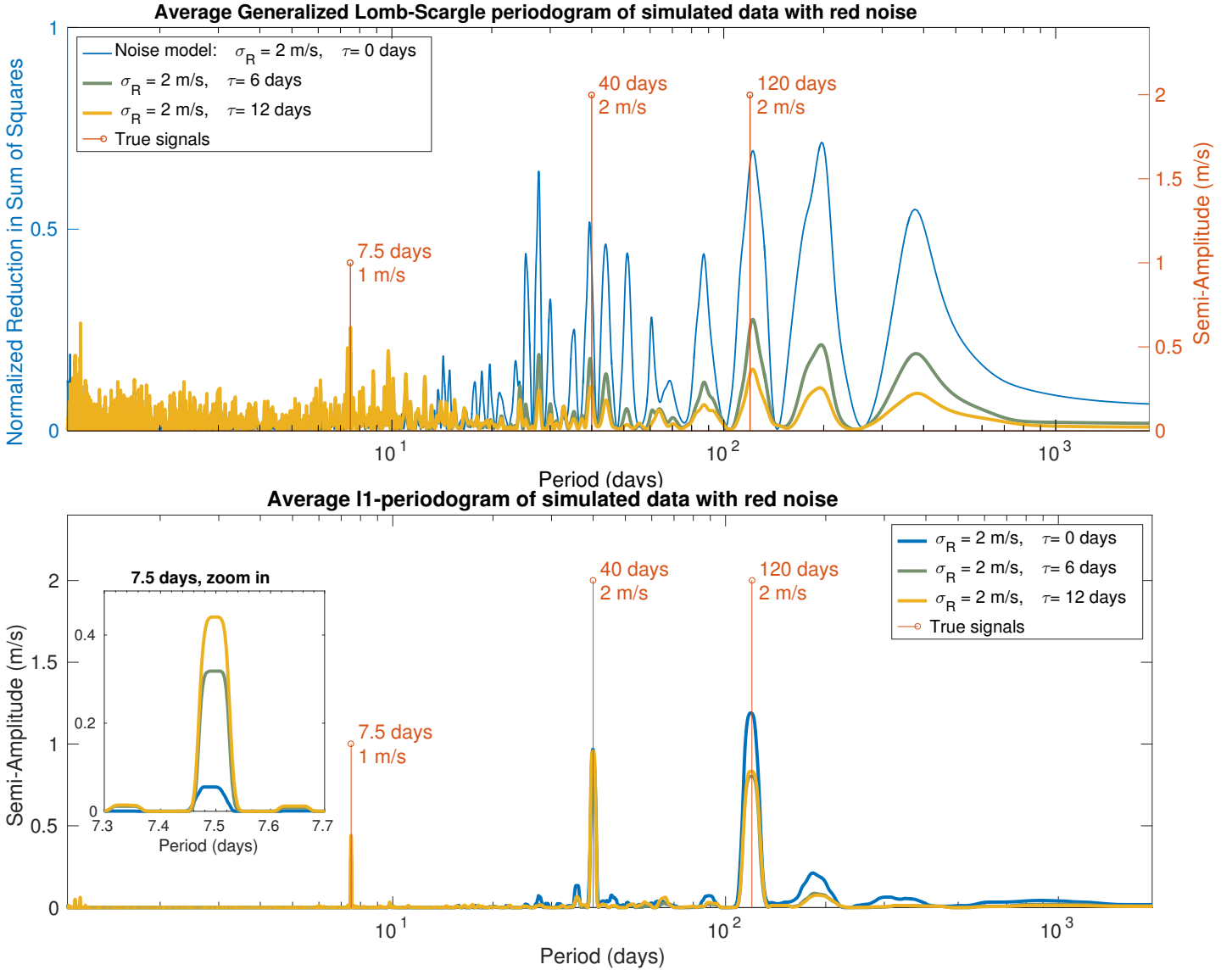
Figure B1. Average ℓ_1-periodogram for 50 data sets generated with red noise of characteristics $\sigma_W = 0, \sigma_R = 2$ m.s^{-1} and $\tau = 12$ days according to model (14). The curves correspond to the solutions of $(11, \mathrm{BP}_{\varepsilon,W})$ with different weight matrices W whose parameters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

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

We compute the number of times the maximum peak 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 is greater than the inverse of the total observation time, that is $\left| 1/P_{1,2,3} - 1/P_{\max} \right| > 1/T_{\mathrm{obs}}$.

Figure 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 days with the periodogram, we find that the strongest peak is spurious in

33 cases out of five hundred 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 haven't 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 on figure C2. First of all in 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.
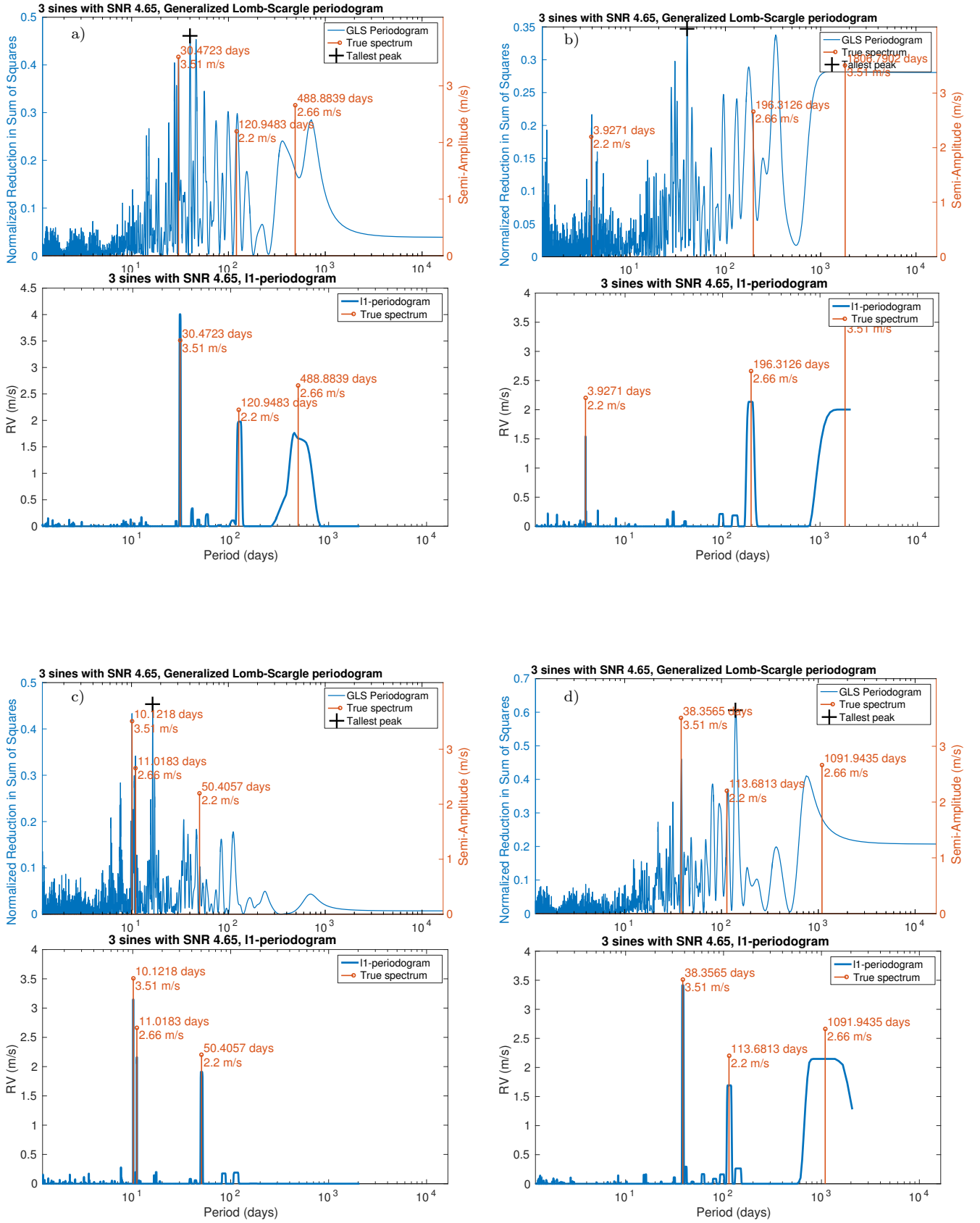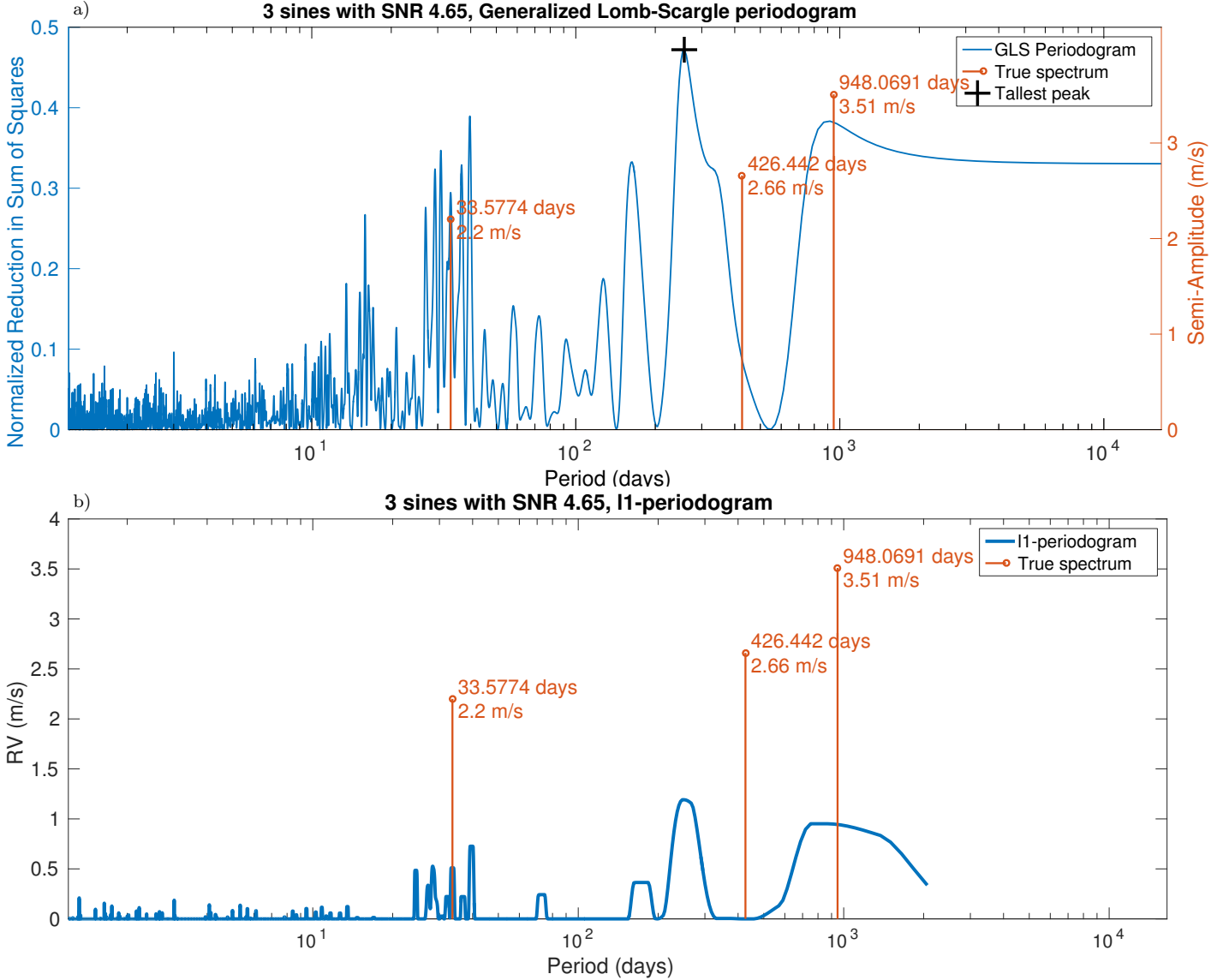
Figure C1. Peak amplitudes and associated FAPs for the four systems analysed

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

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 figure D1).

Let us approximate the error made when fitting an ancillary indicator. We consider the radial velocity signal $y(t) = P(t) + a(t) + \varepsilon(t)$ where $P(t)$ is due to a planetary companion, $a(t)$ is a deterministic signal due to activity and ε is a Gaussian noise of covariance matrix V. We also consider an ancillary measurement $z(t) = a(t) + \varepsilon'$ where $\varepsilon'(t)$ is another

Gaussian noise of covariance matrix V. If we fit $z(t)$ to $y(t)$, we obtain (dropping the t notation):

$$y_{\text{detrend}} = y - y_{\text{fit}} = y - \frac{z^T V^{-1} y}{z^T V^{-1} z} z \qquad (D1)$$

$$y_{\text{detrend}} = y - \frac{(a+\varepsilon')^T V^{-1} (P+a+\varepsilon)^T}{(a+\varepsilon')^T V^{-1} (a+\varepsilon)} (a+\varepsilon'). \qquad (D2)$$

We assume that the noise is small compared to a, which allows to develop the denominator at first order in ε and ε'

$$y_{\text{fit}} \approx \frac{(a+\varepsilon')^T V^{-1} (P+a+\varepsilon)}{a^T V^{-1} a} \left(1 - \frac{\varepsilon'^T V^{-1} a}{a^T V^{-1} a} - \frac{\varepsilon^T V^{-1} a}{a^T V^{-1} a} \right) (a+\varepsilon')$$

After developing that expression at first order in ε 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:

$$\mathbb{E}\{y_{\text{fit}}\} \approx \frac{\sigma^2}{\|a\|^2} P + \tag{D3}$$

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

We would like y_{fit} to be as close to a as possible. This will be better satisfied as the correlation $a^T P$ and as the signal to noise σ^2/a decrease. The fact that a term $a^T P$ appears in the equation above should not be surprising. The mutual coherence defined section 5.2 grasps that the correlation between the parts of the model is an obstacle to 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.

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

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

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

for some constants V_c, V_r, L_c, B_c, B_r. This means that for a given realization (g, 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 g, g'. So according to that model, projecting the radial velocity onto $(\log R'_{HK}, \text{BIS})$ is equivalent to projecting onto (g, 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 $a^T P / \|a\|_{\ell_2}^2$ in equation (D4), assuming the signal $y = P = e^{i\omega t}$ is a pure harmonic of amplitude 1 m/s. Here a designates the FWHM, Bisector span or $\log R'_{HK}$ respectively the red, yellow and purple curves figure 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:

$$\text{Fraction}(\omega) = \frac{(y_\omega - y_{\text{fit}})^T V^{-1} (y_\omega - y_{\text{fit}})}{y_\omega^T V^{-1} y_\omega} \tag{D8}$$

this one is represented by the blue curve figure D2. for the system analysed section 4.6. Only 15% of the energy is absorbed in general, with a maximum of 27% at a period of 2000 days. The peaks at 25 and 12.5 days 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 $\log R'_{hk}$.
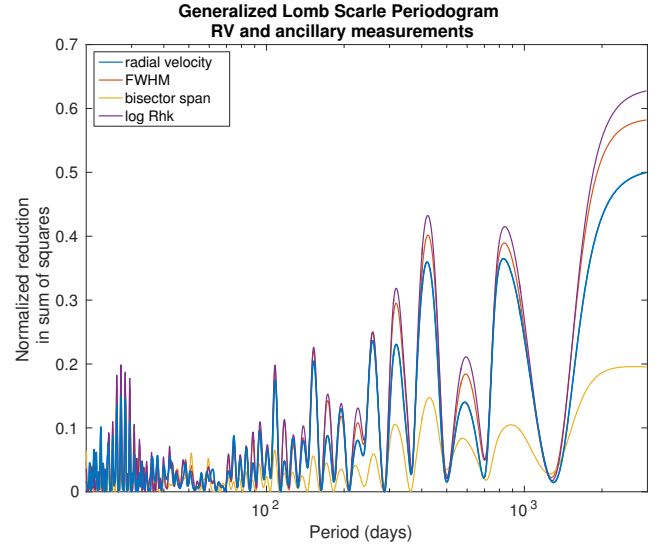
Figure D1. Generalized Lomb-Scargle periodogram of radial velocity and ancillary measurements at low frequencies
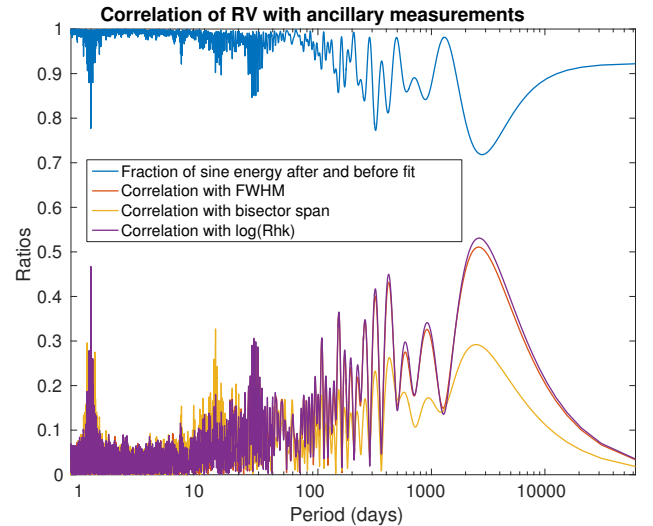
Figure D2. Energy of a cosine function after the fit of the FWHM, bisector span, $\log R'_{hk}$ and a constant

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