SEMI-SUPERVISED STANDARDIZED DETECTION OF PERIODIC SIGNALS WITH APPLICATION TO EXOPLANET DETECTION

S. Sulis,

Aix Marseille Univ., CNRS, CNES, LAM
Marseille, France

D. Mary, L. Bigot

Univ. Côte d'Azur, Obs. de la Côte d'Azur
CNRS, Lagrange UMR 7293, Nice, France

ABSTRACT

We propose a numerical methodology for detecting periodicities in unknown colored noise and for evaluating the 'significance levels' (p-values) of the test statistics. The procedure assumes and leverages the existence of a set of time series obtained under the null hypothesis (a null training sample, NTS) and possibly complementary side information. The test statistic is computed from a standardized periodogram, which is a pointwise division of the periodogram of the series under test to an averaged periodogram obtained from the NTS. The procedure provides accurate p-values estimation through a dedicated Monte Carlo procedure. While the methodology is general, our application is here exoplanet detection. The proposed methods are benchmarked on astrophysical data.

***Index Terms*—** semi-supervised detection, statistical learning, uneven sampling, Monte Carlo, colored noise

1. INTRODUCTION

The detection of periodic signals is perhaps the most visited problem in the signal processing empire [1, 2, 3]. Yet, in the vast forest of detection settings encountered in practical applications, there is a small village of indomitable problems who still holds against theoretical and analytical invaders [4]: these problems combine irregular sampling and unknown noise statistics. Indeed, irregular sampling creates dependencies between periodogram's ordinates (e.g., the Lomb-Scargle periodogram (LSP) [5] or periodograms' variants [6], [7], [8], [9], [10], [11], [12], [13]). These dependencies complicate the statistical interpretation of the periodogram peaks' values and often make analytical derivations of 'significance levels' (p-values) out of reach. The unknown statistics of the colored noise further complicate the picture [14]. Overall, it is hard in this setting to search for constant false alarm rate detectors through pivotal statistics.

This situation is typical in the application framework of the present paper, exoplanet detection in radial velocity (RV) data. In this problem, planets induce a (possibly quasi-) periodic motion to their host star, whose amplitude is proportional to the planets' masses [15].

In the measured RV times series of the host star, the goal is to detect (quasi)-periodic variations. Typically, the sampling grid is irregular (weather-dependent, for instance) and the noise and nuisance parameters are poorly known (they come from numerous perturbations from the host star's and instruments). For these reasons, examples of debated detections are numerous (e.g., planets HD 73256 b [16] [17], AD Leo b [18] [19], Aledebaran c [20] [21], or Kapteyn b and c [22], [23]). We will discuss at the end of this paper the particular case of an Earth-mass planet orbiting α Centauri B, a detection claimed in [24] and controversed in [25, 26, 27].

We consider in this paper a methodology for detecting periodicity in irregularly sampled times series when the noise has partially unknown statistics. The proposed "magic potion" has three ingredients. First, we assume that the detection is semi-supervised, in the sense that some knowledge about the noise is available through a null training sample (NTS) – a few time series of the noise alone – and possibly further ancillary data on the nuisance parameters. Second, we leverage this information to standardize the periodogram [14] from which a test statistic is obtained. Third, we propose a Monte Carlo (MC) procedure to evaluate the p-values of the test statistic.

Periodicity detection approaches using randomized inference exist (e.g., [28], [27]) but they do not consider the case of unknown colored noise. To our knowledge, there is currently no procedure attempting to derive accurate significance levels in the considered setting. The procedure presented below extends previous works, whose results were mostly limited to regular sampling [14], [29], [30] or to situations with no unknown mean signal [31]. The proposed algorithms are available on GitHub[1].

The next section presents the model and the detection approach (Algorithm 1). The third section details the Monte Carlo method (Algorithm 2) designed for evaluating the p-value of a test statistic produced by Algorithm 1. The last Section presents applications of the proposed procedure to RV data of the Sun and α Centauri B.

2. SEMI-SUPERVISED STANDARDIZED DETECTION

Given an irregularly sampled data time series $\boldsymbol{x} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^\top$, we consider the following composite hypothesis testing problem:

$$\begin{cases} \mathcal{H}_0 \ : \boldsymbol{x} = \boldsymbol{d}|\mathcal{M}_d(\boldsymbol{\theta}_d) + \boldsymbol{n}, \\ \mathcal{H}_1 \ : \boldsymbol{x} = \mathbf{s}|\mathcal{M}_s(\boldsymbol{\theta}_s) + \boldsymbol{d}|\mathcal{M}_d(\boldsymbol{\theta}_d) + \boldsymbol{n}, \end{cases} \boldsymbol{n} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \qquad (1)$$

with $\boldsymbol{n}$ a zero mean, Gaussian stochastic noise component of unknown covariance matrix. We assume that $\boldsymbol{n}$ is the sum of a colored Gaussian component with covariance matrix $\boldsymbol{\Sigma}_c$ and a white Gaussian noise component (WGN) with covariance matrix $\sigma_w^2 \boldsymbol{I}$, that is, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_c + \sigma_w^2 \boldsymbol{I}$, with $\boldsymbol{\Sigma}_c$ and σ_w^2 unknown. Notation $\boldsymbol{z}|\mathcal{M}$ means that the vector $\boldsymbol{z}$ is generated from model $\mathcal{M}$. This model can be deterministic (for instance if $\boldsymbol{z}$ can be written $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{\alpha}$ for some parameter matrix $\boldsymbol{A}$ and fixed coefficient vector $\boldsymbol{\alpha}$) or stochastic (for instance if $\boldsymbol{z}$ is a realization of a p^{th} order autoregressive (AR) process, whose parameters are the filter's coefficients). The nuisance signal $\boldsymbol{d}$ (unknown mean under $\mathcal{H}_0$) can have various sources: dark spots or bright plages at the stellar surface, instrumentals drifts, etc. It is customary to consider this trend as a deterministic signature that

[1] https://github.com/ssulis/3SD

is estimated and removed (see e.g., [24]), but nothing prevents from opting for a random description of this term. We assume generically that $\boldsymbol{d}$ is generated from some model $\mathcal{M}_d$ with parameters $\boldsymbol{\theta}_d$. The signal $\boldsymbol{s}$ is an unknown, deterministic, periodic or quasi-periodic signal from some model $\mathcal{M}_s$.

We shall consider the semi-supervised situations where part or all of the following side information is available: 1) For $\boldsymbol{d}$, the model $\mathcal{M}_d(\boldsymbol{\theta}_d)$ is available (but the corresponding parameter $\boldsymbol{\theta}_d$ vector is indeed unknown). 2) For the stochastic part $\boldsymbol{n}$, a set of $L \ll N$ time series of the noise (the NTS) is available:

$$\mathcal{T}_L := \{\boldsymbol{n}^{(i)}\}, i = 1, \ldots, L, \quad \boldsymbol{n}^{(i)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}). \tag{2}$$

To cope with unknown noise correlations, our approach is based on periodogram standardization [14]. We denote the periodogram of a time series $\boldsymbol{x}$ by $\boldsymbol{p}(\boldsymbol{x})$ (a vector concatenating the periodogram values at all frequencies) or by $\boldsymbol{p}(\nu_k|\boldsymbol{x})$ (the ordinate of $\boldsymbol{p}$ at frequency ν_k). Frequencies (in number N_f) are computed in a set denoted by Ω (this set is determined by the user). A reference, averaged periodogram can be obtained from a set $\mathcal{T}_L$ of L times series $\boldsymbol{x}_\ell$ as

$$\overline{\boldsymbol{p}}_L(\nu_k|\mathcal{T}_L) := \frac{1}{L}\sum_{\ell=1}^{L} \boldsymbol{p}_\ell(\nu_k|\boldsymbol{x}^{(\ell)}), \tag{3}$$

also written $\overline{\boldsymbol{p}}_L(\mathcal{T}_L)$ for short. The standardized periodogram is defined as

$$\widetilde{\boldsymbol{p}}(\nu_k|\boldsymbol{x}, \mathcal{T}_L) := \frac{\boldsymbol{p}(\nu_k)}{\overline{\boldsymbol{p}}_L(\nu_k|\mathcal{T}_L)}, \tag{4}$$

also written $\widetilde{\boldsymbol{p}}(\boldsymbol{x}, \mathcal{T}_L)$ for short.

In the approach above, the considered type of periodogram P (see Sec. 1) is chosen by the user. Once the periodogram $\widetilde{\boldsymbol{p}}$ has been computed, it remains to choose a test, denoted by T. If information on the signal $\boldsymbol{s}$ is available it can be exploited in the test through, for instance, a Generalized Likelihood Ratio approach. However, Maximum Likelihood Estimation under both hypotheses may be intractable owing to the rather large number of unknown parameters (this is the case for the Keplerian signatures of exoplanets). Hence, it may be useful to use other test statistics [14], [32], [33]. Here, in order to instantiate our approach, we will use the classical and frequently used 'Max-test', whose test statistic writes:

$$T_M(\widetilde{\boldsymbol{p}}) := \max_k \widetilde{\boldsymbol{p}}(\nu_k). \tag{5}$$

Computing (5) leads to an observed test statistic, say t_M. A crucial question in practice is then: How 'significant' is t_M ? Often, an answer is proposed in the form of a p-value, denoted by v below:

$$\mathsf{v}(t_M) := \Pr(T_M > t_M \mid \mathcal{H}_0) \tag{6}$$

so that $\mathsf{v}(t_M) = 1 - \Phi_{T_M}(t_M)$ with Φ_{T_M} the Cumulative Distribution Function (CDF) of T_M. Unfortunately, in the situation of model (1), the distribution of T_M under both hypotheses is unknown, since it depends on several unknown parameters. Before turning to procedures aimed at evaluating the p-values of the considered semi-supervised detection procedure, we summarize this procedure in Algorithm 1. Given a (possibly unevenly sampled) time series under test $\boldsymbol{x}$, Algorithm 1 applies test T to a standardized periodogram computed with periodogram type P and outputs a test statistics $t(\boldsymbol{x})$. If a nuisance signal $\boldsymbol{d}|\mathcal{M}_d$ is assumed present in the time series, row 2 estimates the parameters of this model. This leads to parameter estimates $\widehat{\boldsymbol{\theta}}_d$, from which an estimate of the nuisance signal $\widehat{\boldsymbol{d}}|\mathcal{M}_d(\widehat{\boldsymbol{\theta}}_d)$ is generated and subtracted from the time series (row 3). If available, the NTS is incorporated in the standardization (rows 6-7). If not, the noise is considered white with unknown variance and standardization amounts to a simple global scaling (rows 8-10).

Algorithm 1: Considered standardized detection procedure. The procedure is semi-supervised if side information $\mathcal{T}_L$ or $\mathcal{M}_d$ is available.

Inputs : $\boldsymbol{x}$: Times series under test
(P,T): selected couple (periodogram, test)
$\boldsymbol{\Omega}$: considered set of frequencies
$\mathcal{T}_L$ and/or $\mathcal{M}_d$

Output: Test statistic $t(\boldsymbol{x})$

1 **if** $\mathcal{M}_d \neq \emptyset$ **then**
2 | Estimate $\widehat{\boldsymbol{\theta}}_d$
3 | $\boldsymbol{x} \leftarrow \boldsymbol{x} - \widehat{\boldsymbol{d}}|\mathcal{M}_d(\widehat{\boldsymbol{\theta}}_d)$
4 **end**
5 $\boldsymbol{p}(\boldsymbol{x}) \leftarrow$ Apply P to $\boldsymbol{x}$
6 **if** $\mathcal{T}_L \neq \emptyset$ **then**
7 | Compute $\overline{\boldsymbol{p}}_L(\mathcal{T}_L)$ as in (3)
8 **else**
9 | $\widehat{\sigma^2} \leftarrow$ Estimate var($\boldsymbol{x}$)
10 | $\overline{\boldsymbol{p}}_L \leftarrow \widehat{\sigma^2}\boldsymbol{1}$
11 **end**
12 Compute $\widetilde{\boldsymbol{p}}$ as in (4)
13 $t(\boldsymbol{x}) \leftarrow$ Apply T to $\widetilde{\boldsymbol{p}}$

3. ESTIMATION OF P-VALUES

3.1. Basic approach

When unknown nuisance parameters are present under $\mathcal{H}_0$, the computation of p-values is not straightforward. Several approaches exist [34]. A simple approach allowing to account for uncertainties in the model's parameters considers the parameters as random and computes the expectation of p-values over a prior parameter's distribution (prior predictive p-values, [35]). This is the approach of the MC procedure proposed in Algorithm 2. This algorithm samples a large number of test statistics that are consistent with the data and the models, by randomly perturbing the parameters within their uncertainties according to some prior distribution. This allows to estimate an expected p-value for any test statistic t produced by Algorithm1.

Algorithm 2 covers the most general situation where both $\mathcal{T}_L \neq \emptyset$ and $\mathcal{M}_d \neq \emptyset$. In addition to the parameters of Algorithm 1, Algorithm 2 requires the following input information.
If $\mathcal{T}_L \neq \emptyset$, a model $\mathcal{M}_n$ is required to generate synthetic samples of the NTS. This model involves parameters that are estimated from $\mathcal{T}_L$ and called $\widehat{\boldsymbol{\theta}}_n|\mathcal{M}_n$. For instance, in our application, $\mathcal{M}_n$ can be an AR process and $\widehat{\boldsymbol{\theta}}_n|\mathcal{M}_n$ are the AR order and filter coefficients.
If $\mathcal{T}_L = \emptyset$, an estimate of the variance of the WGN component ($\widehat{\sigma^2}_w$) is required, along with a perturbation prior, which is typically obtained from error estimates. Sample noise time series are then composed by i.i.d. Gaussian samples with variance $\widehat{\sigma^2}_w + \epsilon_w$, where ϵ_w is typically drawn from a prior distribution (denoted by π) with zero mean and scale parameter denoted by Δ_w. In the simulations, we considered uniform and Gaussian priors and noticed negligible dependence on the prior's shape and scale parameter.
If $\mathcal{M}_d \neq \emptyset$, Algorithm 2 requires the model and the corresponding parameters $\widehat{\boldsymbol{\theta}}_d|\mathcal{M}_d$. Those can be obtained from ancillary measurements and/or from $\boldsymbol{x}$. In the second application case of Sec. 4 for instance, the nuisance signal can be written as a linear combination of low order polynomials, sinusoids of known frequencies and deterministic signals obtained from ancillary measurements. For the MC sampling, the estimated parameters $\widehat{\boldsymbol{\theta}}_d|\mathcal{M}_d$ are perturbed according to a prior distribution denoted by $\pi(\boldsymbol{0}, \boldsymbol{\Delta}_d)$.

Algorithm 2: Monte Carlo procedure for estimating the p-value of the result of Algorithm 1 along with confidence intervals.

Inputs : $\boldsymbol{x}$: Times series under test
(P,T): selected couple (periodogram, test)
$\boldsymbol{\Omega}$: considered set of frequencies
b, B: Monte Carlo sample size
π: parameters' prior distribution
if $\mathcal{T}_L \neq \emptyset$ **then**
  $\mathcal{M}_n$: parametric model for $\boldsymbol{n}$
  $\widehat{\boldsymbol{\theta}}_n|\mathcal{M}_n, \mathcal{T}_L$: estimated parameters
else
  $\widehat{\sigma^2}_w$: estimated variance of WGN
  $\boldsymbol{\Delta}_w$: scale parameter for prior π on $\widehat{\sigma^2}_w$
end
if $\mathcal{M}_d \neq \emptyset$ **then**
  $\widehat{\boldsymbol{\theta}}_d|\mathcal{M}_d$: estimated parameters
  $\boldsymbol{\Delta}_d$: scale parameters for prior π on $\widehat{\boldsymbol{\theta}}_d$
end

Output: $\widehat{\mathsf{v}}(t)$ and 90% confidence interval

1 **for** $i = 1, \ldots, B$ **do**
2 **if** $\mathcal{T}_L \neq \emptyset$ **then**
3 $\mathcal{T}_L^{(i)} \leftarrow$ Generate from $\mathcal{M}_n(\widehat{\boldsymbol{\theta}}_n)$
4 $\widehat{\widehat{\boldsymbol{\theta}}}_n^{(i)} \leftarrow$ Estimate from $\mathcal{T}_L^{(i)}|\mathcal{M}_n$
5 **end**
6 **for** $j = 1, \ldots, b$ **do**
7 **if** $\mathcal{T}_L \neq \emptyset$ **then**
8 $\mathcal{T}_L^{(i,j)} \leftarrow \{\boldsymbol{n}^{(i,j,\ell)}|\widehat{\widehat{\boldsymbol{\theta}}}_n^{(i)}\}_{\ell=1,\ldots,L}$
9 $\boldsymbol{x}^{(i,j)} \leftarrow \boldsymbol{n}^{(i,j,L+1)}|\widehat{\widehat{\boldsymbol{\theta}}}_n^{(i)}$
10 **else**
11 $\epsilon_w \leftarrow$ Generate from $\pi(0, \boldsymbol{\Delta}_w)$
12 $\widehat{\widehat{\sigma^2}}_w^{(i,j)} \leftarrow \widehat{\sigma^2}_w + \epsilon_w$
13 $\boldsymbol{w}^{(i,j)} \sim \{\mathcal{N}(\boldsymbol{0}, \widehat{\widehat{\sigma^2}}_w^{(i,j)})$
14 $\boldsymbol{x}^{(i,j)} \leftarrow \boldsymbol{w}^{(i,j)}$
15 **end**
16 **if** $\mathcal{M}_d \neq \emptyset$ **then**
17 $\boldsymbol{\epsilon}^{(i,j)} \leftarrow$ Generate from $\pi(\boldsymbol{0}, \boldsymbol{\Delta}_d)$
18 $\widehat{\boldsymbol{d}}^{(i,j)} \leftarrow$ Generate $\mathcal{M}_d(\widehat{\boldsymbol{\theta}}_d + \boldsymbol{\epsilon}^{(i,j)})$
19 $\boldsymbol{x}^{(i,j)} \leftarrow \boldsymbol{x}^{(i,j)} + \widehat{\boldsymbol{d}}^{(i,j)}$
20 **end**
21 $t^{(i,j)} =$ Algorithm 1 $\left(\boldsymbol{x}^{(i,j)}, (\mathsf{P}, \mathsf{T}), \boldsymbol{\Omega}, \mathcal{T}_L^{(i,j)}, \mathcal{M}_d\right)$
22 **end**
23 $\widehat{\Phi}_T^{(i)} \leftarrow$ Estimate CDF from the $\{t^{(i,j)}\}_{j=1,\ldots,b}$
24 $\widehat{\mathsf{v}}^{(i)}(t) \leftarrow 1 - \widehat{\Phi}_T^{(i)}(t)$
25 **end**
26 $\widehat{\mathsf{v}}(t) \leftarrow \frac{1}{B}\sum_{i=1}^{B} \widehat{\mathsf{v}}^{(i)}(t)$
27 90% confidence interval $\leftarrow \{\widehat{\mathsf{v}}^{(i)}(t)\}_{i=1,\ldots,B}$

The procedure starts with a loop aimed at generating B fake NTS (row 1). To this end, the input parameters are used to generate a fake NTS $\mathcal{T}_L^{(i)}$ (row 3) from which the model parameters are re-estimated, providing perturbed parameters (row 4). Note that no prior is required here.

The algorithm enters a second loop (row 6) aimed at generating b fake times series under test $\{x^{(i,j)}\}$, $j = 1, \ldots, b$, each containing a different perturbed version of a colored noise signal (row 9), a WGN component (row 14) and a nuisance signal (row 19). To this end, a fake NTS $\mathcal{T}_L^{(i,j)}$ (row 8) and a fake time series $\boldsymbol{n}^{(i,j,L+1)}$ (row 9) are computed from $\widehat{\widehat{\boldsymbol{\theta}}}_n^{(i)}$. The nuisance signal $\widehat{\boldsymbol{d}}^{(i,j)}$ is obtained as described above (rows 17-18) and added to the fake time series (row 19). Note that if $\mathcal{T}_L \neq \emptyset$, the WGN component does not need to be added as it is generated in the NTS. Algorithm 1 is then applied to the resulting synthetic time series and NTS, with the same parameters that were used in Algorithm . The main loop leads to B sets of empirical CDF and p-values (rows 23-24). The average p-value and a confidence interval are finally computed (rows 26-27).

3.2. Model errors and Algorithm 2

In Algorithm 2, the MC samples are generated under the same model (1) that is used in the semi-supervised detection (Algorithm 1). In practice, however, model (1) is never exact so that the p-values estimated by Algorithm 2 may depart from the p-values that would be obtained by repeatedly running Algorithm 1 under the true (not modeled) null hypothesis. This effect can be illustrated analytically in the case of regular sampling. Consider a time series of N points composed of WGN of known variance under $\mathcal{H}_0$. The classical periodogram is then distributed when properly scaled as a χ^2 random variable with two degrees of freedom and non-centrality parameter $\lambda = 0$: $\frac{2}{\sigma^2}\boldsymbol{p}(\nu_k) \sim \chi^2_{2,\lambda=0}$, $\forall \nu_k \in \Omega$ ([2], Th. 6.2). Hence, the p-value attributed to an observed test statistic t of the Max-test is

$$\widehat{\mathsf{v}}(t) = \Pr(T_M > t|\mathcal{H}_0) = 1 - \Pr(\forall k, \boldsymbol{p}(\nu_k) \leq t) = 1 - \Phi^{N_f}_{\chi^2_{2,0}}(t), \quad (7)$$

where $\Phi_{\chi^2_{2,\lambda}}(t)$ is the CDF of a $\chi^2_{2,\lambda}$ random variable. Consider now the case where this model is erroneous because some frequencies are actually affected by a nuisance signal under $\mathcal{H}_0$. In the simplest case where this signal affects a single frequency (the error model then amounts to an ignored residual sinusoid in the data under $\mathcal{H}_0$, with amplitude, say A, and frequency on the Fourier grid), one can show that $\frac{2}{\sigma^2}\boldsymbol{p}(\nu_k) \sim \chi^2_{2,\lambda=\frac{nA^2}{2\sigma^2}}$: at the nuisance signal frequency, the scaled periodogram is now a non-central $\chi^2_{2,\lambda}$ random variable and the distribution at other frequencies remains unchanged. Hence, the true p-value is in fact

$$\mathsf{v}(t) = 1 - \Phi_{\chi^2_{2,\frac{NA^2}{2\sigma^2}}}(t) \cdot \Phi^{(N_f-1)}_{\chi^2_{2,0}}(t). \quad (8)$$

The CDF of a χ^2 random variable is related to the generalized Marcum Q-function Q_1 by $1 - \Phi_{\chi^2_2(t,\frac{NA^2}{2\sigma^2})} = Q_1(\sqrt{\frac{N}{2}}\frac{A}{\sigma}, \sqrt{t})$ and Q_1 is an increasing function of the non centrality parameter $\sqrt{\frac{N}{2}}\frac{A}{\sigma}$ [36]. It can be easily checked that this implies that $\forall A > 0$, $\mathsf{v}(t) > \widehat{\mathsf{v}}(t)$. In other words, in presence of model errors, true p-values may be larger than estimated under no model error, an effect that tends to artificially increase the number of possible discoveries. For instance, with $N = 100$, $A = \sigma^2/2 = 1/2$, $N_f = 50$ and $t = 17$, $\widehat{\mathsf{v}}(t) = 1\%$ while $\mathsf{v}(t) = 33\%$. Indeed, the precise characterization of model errors and their impact on p-values is by definition out of reach in general. Irregular sampling and unknown colored noise also make the situation far more complicated than in this toy (yet illustrative) example. In such situations, an interesting added value of Algorithm 2 is to allow for studying the robustness of a specific detection procedure to model errors by sampling from some model and conducting the detection and p-value estimation procedures with another model. We show an example in the next Section.

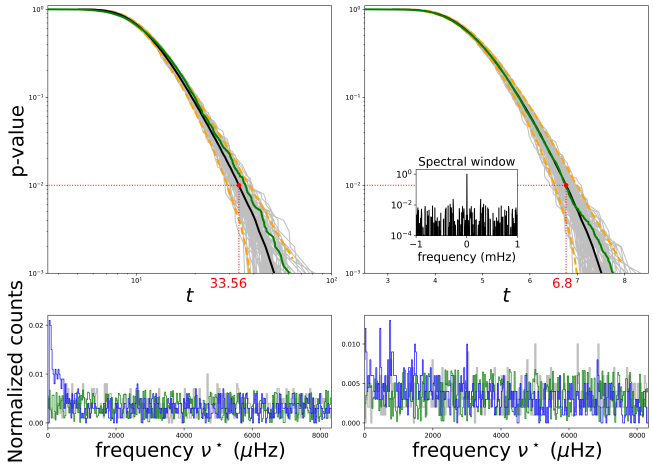

Fig. 1: Accuracy of Algorithm 2 in estimating the p-value of the test statistic produced by Algorithm 1 (see text in Sec. 4).

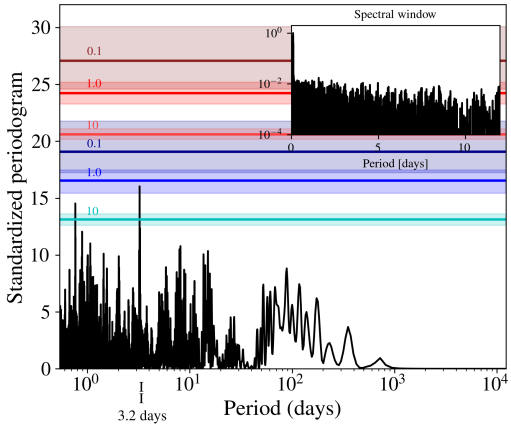

Fig. 2: Impact of model errors. Black: data PLS . Blue: prior predictive p-values levels 0.1, 1 and 10% (solid) with their 90% confidence intervals (shade) estimated assuming model $\mathcal{M}_d$ is true. Red: True p-value levels if model $\mathcal{M}'_d \neq \mathcal{M}_d$ is true. Inset: spectral window.

4. APPLICATION TO EXOPLANET DETECTION

We first investigate how accurately Algorithm 2 allows to estimate the p-values. The time series under test has $N = 288$ points drawn randomly in an interval of two days. The data come from the GOLF instrument onboard the SoHO spacecraft. The deterministic term (solar activity and instrumental trends) has been filtered out before data publication, hence $\boldsymbol{d} = \mathbf{0}$ and $\mathcal{M}_d = \emptyset$ here. An NTS is used in the semi-supervised detection Algorithm 1, with $L = 5$ time series, which were obtained through magneto-hydrodynamic (MHD) simulations of the solar atmosphere [37] and instrumental WGN estimated from the instrument (see [30] for details). In Algorithm 2, $\mathcal{M}_n$ is an AR model, whose order was initially selected using Akaike's Final Prediction Error criterion [38]. The filter coefficients are obtained through the standard Yule-Walker equations [39].

Note that our NTS does not exactly undergo the noise model, since MHD simulations are not genuine stellar noise time series. This is an additional difficulty for validating the approach, but this is worth it since if we show that MHD simulations can transparently be used in place of true noise series for standardization, the proposed MHD based approach could virtually be applied to planet search around any type of star.

Fortunately for the purpose of validating Algorithm 2, GOLF instrument provides a 25-year-long time series of the Sun, so this data set can be used as ground truth (called Oracle below): indeed, Algorithm 1 can be run on a large number (14200) of noise time series by using real (not MHD) noise series for standardization. (Note that in the considered frequency range, solar system's planetary signatures are indeed not present.) The resulting p-values provide a benchmark against which the p-values of Algorithm 2 can be compared.

In the left (resp. right) panels of Fig. 1, the couple (P, T) corresponds to the Lomb-Scargle periodogram (LSP) and Max test (resp. Generalized LSP [8] and test T_C). Test T_C uses the C^{th} largest value of $\widetilde{\boldsymbol{p}}$ as test statistic [40, 41]. This test can be more powerful than the Max test for signatures that are not very sparse in the Fourier domain (typically C is set in the range $[1, 20]$) [29, Chap. 3.5.1]. The top row shows $B = 100$ estimates $\widehat{\mathsf{v}}^{(i)}(t)$ produced by Algorithm 2 with $b = 10^3$ (grey), the average $\widehat{\mathsf{v}}(t)$ (black) and the true (Oracle) p-value (green). The orange beam is the 90% envelope. For instance, for $t = 33.56$, $\widehat{\mathsf{v}} = 1\%$ with a 90% 'confidence interval' of $[0.4, 1.6]\%$, while the true p-value is 1.3%. Both panels show that Algorithm 2 evaluates p-values quite accurately with perhaps a slight bias of $\widehat{\mathsf{v}}(t)$ at large t due to the limited number of MC simulations.

The bottom panels show the empirical distributions of $b = 10^3$ frequency indices $\nu^* := \arg\max_{\nu_k} \widetilde{\boldsymbol{p}}(\nu_k)$ found respectively when running Algorithm 2 (grey), the Oracle (green, only contours are shown for visibility) and a classical, un-standardized detection procedure where the Max test is applied to $\boldsymbol{p}$ instead of $\widetilde{\boldsymbol{p}}$ in Algorithms 1 and 2 (blue). We see that while the standardized approaches lead to a uniform distribution of ν^*, this is not the case of the un-standardized approach, for which ν^* hits more frequently low frequencies (because the noise has higher power spectrum density there). This shows that the considered standardization properly mitigates the adversarial effects of the partially unknown colored noise. Other results (not shown) indicate that the conclusions above remain valid for other kinds of sampling grids. In particular, Algorithm 2 precisely recovers the analytical formula of the p-value from [14] for regular sampling.

We turn to robustness against model errors. Fig 2 shows the LS periodogram of 4-year-long RV data taken from HARPS spectrograph of the star α Centauri B [24]. Here, Algorithm 1 was run with a noise model as in [24]: $\mathcal{M}_d$ is a linear model with 23 parameters and $\mathcal{T}_L = \emptyset$. Using permutation bootstrap, the p-value of the largest peak was estimated as 0.02% in [24]. The results of our Algorithm 2 with model $\mathcal{M}_d$ are displayed in blue: accounting for estimation errors in the nuisance parameters leads to a higher p-value of about 1.5%. There is some uncertainty in the model assumed by [24], however: a similar though different model (say, $\mathcal{M}'_d$), based on Gaussian processes was proposed in [26]. It is therefore interesting to investigate how Algorithm 1 run with model $\mathcal{M}_d$ reacts when the data in reality undergo the slightly different model $\mathcal{M}'_d$. To this end, Algorithm 2 was run with $\mathcal{M}'_d$, except in row 21, which used model $\mathcal{M}_d$. The results are shown in Fig. 2, red levels. The p-value of the 3.2d peak, now 63.43%, indicates that this peak's height is in fact quite ordinary under the null hypothesis with model $\mathcal{M}'_d$, showing that the 3.2d detection is not robust under such model errors.

In conclusion, we have proposed a semi-supervised detection procedure and a MC method allowing to evaluate the resulting p-values and the impact of some model error. This procedure is designed to leverage ancillary data such as a NTS, that is used for periodogram standardization. The method is quite versatile in the periodograms, tests and noise models that can be plugged-in. We have illustrated and validated the procedure for exoplanet detection but it can be easily adapted to other periodicity detection problems.

5. REFERENCES

[1] A. Schuster, "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena," *Terrestrial Magn.*, vol. 3, no. 1, pp. 13–41, 1898.

[2] T.H. Li, *Time series with mixed spectra*, CRC Press, 2014.

[3] S.M. Kay, *Fundamentals of Statistical signal processing. Vol II : Detection theory.*, Prentice-Hall, Inc, 1998.

[4] A. Uderzo and R. Goscinny, *Asterix le gaulois*, 1961.

[5] J. D. Scargle, "Studies in astronomical time series analysis.," *ApJ*, vol. 263, pp. 835–853, 1982.

[6] S. Ferraz-Mello, "Estimation of Periods from Unequally Spaced Observations," *AJ*, vol. 86, pp. 619, 1981.

[7] A. Cumming, "Detectability of extrasolar planets in radial velocity surveys," *MNRAS*, vol. 354, no. 4, pp. 1165–1176, 2004.

[8] M. Zechmeister and M. Kürster, "The generalised LS periodogram. A new formalism for the floating-mean and Keplerian periodograms," *A&A*, vol. 496, no. 2, pp. 577–584, 2009.

[9] R. V. Baluev, "PlanetPack: A RV time-series analysis tool facilitating exoplanets detection, characterization, and dynamical simulations," *Astron. Comput*, vol. 2, pp. 18–26, 2013.

[10] J. S. Jenkins et al., "Improved signal detection algorithms for unevenly sampled data. Six signals in the radial velocity data for GJ876," *MNRAS*, vol. 441, no. 3, pp. 2253–2265, 2014.

[11] M. Tuomi et al., "Bayesian search for low-mass planets around nearby M dwarfs - estimates for occurrence rate based on global detectability statistics," *MNRAS*, vol. 441, no. 2, pp. 1545–1569, 2014.

[12] N. Hara et al., "RV data analysis with compressed sensing techniques," *MNRAS*, vol. 464, no. 1, pp. 1220–46, 2017.

[13] N. C. Hara et al., "Testing whether a signal is strictly periodic. Application to disentangling planets and stellar activity in radial velocities," p. arXiv:2106.01365, 2021.

[14] S. Sulis, D. Mary, and L. Bigot, "A Study of Periodograms Standardized Using Training Datasets and Application to Exoplanet Detection," *IEEE Trans. on Signal Processing*, vol. 65, pp. 2136–2150, 2017.

[15] M. Perryman, *The Exoplanet Handbook*, Cambridge University Press, 2011.

[16] S. Udry et al., "The CORALIE survey for southern extra-solar planets. X. A Hot Jupiter orbiting HD 73256," *A&A*, vol. 407, pp. 679–684, 2003.

[17] K. Ment et al., "RV from the N2K Project: Six New Cold Gas Giant Planets Orbiting HD 55696, HD 98736, HD 148164, HD 203473, and HD 211810," *ApJ*, vol. 156, no. 5, pp. 213, 2018.

[18] M. Tuomi et al., "AD Leonis: RV Signal of Stellar Rotation or Spin-Orbit Resonance?," *ApJ*, vol. 155, no. 5, pp. 192, 2018.

[19] I. Carleo et al., "The GAPS Programme at TNG. XXI. A GIARPS case study of known young planetary candidates: confirmation of HD 285507 b and refutation of AD Leonis b," *A&A*, vol. 638, pp. A5, 2020.

[20] A. P. Hatzes et al., "Long-lived, long-period radial velocity variations in Aldebaran: A planetary companion and stellar activity," *A&A*, vol. 580, pp. A31, 2015.

[21] K. Reichert et al., "Precise radial velocities of giant stars. XII. Evidence against the proposed planet Aldebaran b," *A&A*, vol. 625, pp. A22, 2019.

[22] G. Anglada-Escude et al., "Two planets around Kapteyn's star: a cold and a temperate super-Earth orbiting the nearest halo red dwarf.," *MNRAS*, vol. 443, pp. L89–L93, 2014.

[23] A. Bortle et al., "A Gaussian Process Regression Reveals No Evidence for Planets Orbiting Kapteyn's Star," *ApJ*, vol. 161, no. 5, pp. 230, 2021.

[24] X. Dumusque et al., "An Earth-mass planet orbiting α Centauri B," *Nature*, vol. 491, no. 7423, pp. 207–211, 2012.

[25] A. P. Hatzes, "The Radial Velocity Detection of Earth-mass Planets in the Presence of Activity Noise: The Case of α Centauri Bb," *ApJ*, vol. 770, no. 2, pp. 133, 2013.

[26] V. Rajpaul et al., "Ghost in the time series: no planet for Alpha Cen B," *MNRAS*, vol. 456, no. 1, pp. L6–L10, 2016.

[27] P. Toulis and J. Bean, "Randomization Inference of Periodicity in Unequally Spaced Time Series with Application to Exoplanet Detection," *arXiv e-prints*, p. arXiv:2105.14222, 2021.

[28] M. Süveges et al., "A comparative study of four significance measures for periodicity detection in astronomical surveys," *MNRAS*, vol. 450, no. 2, pp. 2052–2066, 2015.

[29] S. Sulis, *Statistical methods using hydrodynamic simulations of stellar atmospheres for detecting exoplanets in radial velocity data*, Theses, Université Côte d'Azur, 2017.

[30] Sulis, S., Mary, D., and Bigot, L., "3d magneto-hydrodynamical simulations of stellar convective noise for improved exoplanet detection - i. case of regularly sampled radial velocity observations," *A&A*, vol. 635, pp. A146, 2020.

[31] S. Sulis, D. Mary, and L. Bigot, "A bootstrap method for sinusoid detection in colored noise and uneven sampling. application to exoplanet detection," in *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 1095–1099.

[32] R.H. Berk and D.H. Jones, "Goodness-of-fit test statistics that dominate the Kolmogorov statistics," *Z WAHRSCHEINLICHKEIT*, vol. 47, no. 1, pp. 47–59, 1979.

[33] D. Donoho and J. Jin, "Higher criticism for detecting sparse heterogeneous mixtures," *Annals of Statistics*, 2004.

[34] M. J. Bayarri and J. O. Berger, "P values for composite null models," *Journal of the American Statistical Association*, vol. 95, no. 452, pp. 1127–1142, 2000.

[35] G. E. P. Box, "Sampling and bayes' inference in scientific modelling and robustness," *Journal of the Royal Statistical Society. Series A (General)*, vol. 143, no. 4, pp. 383–430, 1980.

[36] Y. Sun, Barics A., and S. Zhou, "On the monotonicity, log-concavity and tight bounds of the generalized Marcum and Nutall Q-functions," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1166–1186, 2010.

[37] Å. Nordlund et al., "Solar Surface Convection," *Living Reviews in Solar Physics*, vol. 6, pp. 2, 2009.

[38] H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst. Stat. Math.*, vol. 21, no. 1, pp. 243–247, 1969.

[39] P.J. Brockwell and R.A. Davis, *Time series : theory and methods*, Springer, 1991.

[40] M. Shimshoni, "On Fisher's Test of Significance in Harmonic Analysis," *Geophysical J.*, vol. 23, no. 4, pp. 373–377, 1971.

[41] S-T. Chiu, "Detecting periodic components in a white gaussian time series," *J. R. Stat.*, vol. 51, no. 2, pp. 249–59, 1989.