How reliable is an extrasolar planet detection claim ?
An efficient approach with statistical control of the detection significance.

Sophia Sulis, David Mary, Lionel Bigot

Laboratoire Lagrange, Université Sophia-Antipolis, Observatoire de la Côte d'Azur, Nice, France

Sophia.Sulis@oca.eu, David.Mary@unice.fr, Lionel.Bigot@oca.eu

Abstract

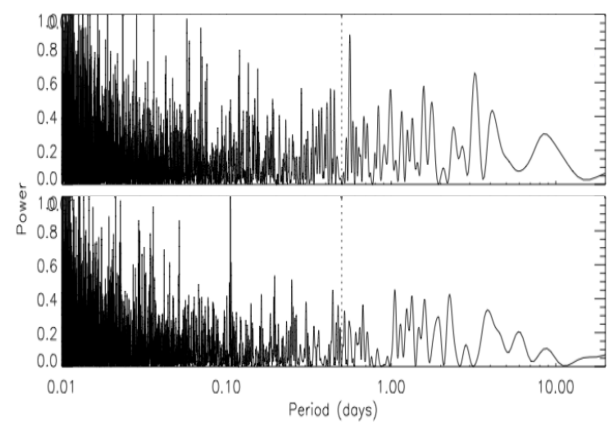

Figure 1: Periodogram of realistic 3D stellar noise simulations at two different epochs. The variations are due to (cf.[1]) :
- Estimation noise (the variance of the periodogram is the DSP itself),
- The stochastic effect of the turbulent convection at the surface,
- The stellar oscillations (at short periods ∼ 5min).

The detection of periodic signals due to Earth-like planets is challenging since it is of the same order (a few cm/s) than the stellar 'noise'. This 'noise' is due to the activity of the star and can have various origins and characteristic time scales : spots, granulation, supergranulation (etc..) and is the main limitation of the detection. In the context of future projects dedicated to telluric planet detection using radial velocities like ESPRESSO (2017) or later the follow up of PLATO (2024), it is mandatory to characterize this noise and developed efficient tools to accurately quantify the probability of detection. This is the aim of the present work. This noise is a stochastic process, with poorly known and star-dependent statistics. Its power is expected to dominate that of the faintest targeted extrasolar signatures. This new adversarial random process challenges the relevance of traditional parametric models used in current detection approaches - in particular their reliability in terms of statistical significance.

In parallel, state-of-the-art numerical hydrodynamic (HD) simulations suggest that it is not unrealistic to assume that such numerical approaches will soon be able to reproduce reliable simulations of the stellar noise (*cf.*[1]). In this contribution, we assume that we dispose of such high precision HD simulations and we pose the following questions : How can we use the output of such codes to devise efficient and reliable detection tests ? What would be the ultimate detection performances when such codes are available ? We propose such a detection method and we analyze its performances. We show in particular that this non parametric approach would allow an accurate control of the statistical significance of the detection claims. The test is also asymptotically (in the data points) independent of the stellar noise.

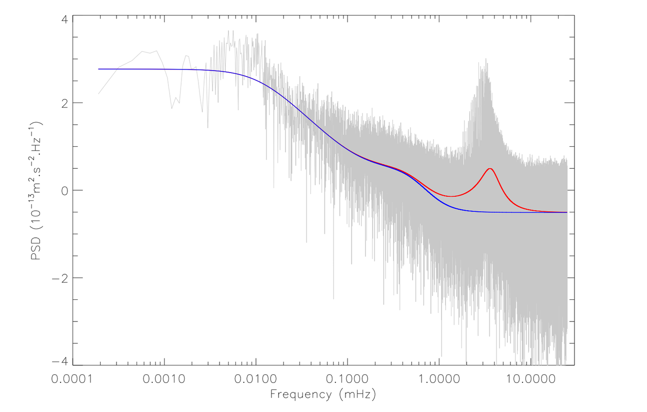

Figure 2: Illustration of the solar power spectral density (DSP) (cf.[2]) estimated over 3 months of regularly sampled observations of the GOLF data on board of the SOHO spacecraft with $\delta t = 20$s.
The red (resp. blue) curves are the results of fits with (resp. without) the stellar oscillation modes (3.5 to 5.5 mHz).

Objectives

- Control (statistically) the **stellar noise** to evaluate **the reliability** associated to a claimed detection of **telluric planets,**
- **Develop detection algorithms** for massive times series dedicated to the case of weak signals buried in partially unknown **colored noise**.

Working hypotheses

- The DSP of the noise (star + detector) is partially known : we assume we can reliably reproduce (through HD simulations) realizations of the stellar noise (*cf.*[1]),
- Large number of data points N and regular sampling (asymptotic independence of the periodogram ordinates at different frequencies).

Periodogram as a Frequency Analysis Tool

- To look for periodicities within the uniformly sampled data, we use the Schuster's periodogram (*cf.*[3]), which is a basis for the majority of the existing methods for detecting hidden periodicities:

$$P(\nu_k) := \frac{1}{N}\Big|\sum_{j=1}^{N} X(t_j)e^{-i2\pi\nu_k t_j}\Big|^2 \quad \text{with } k = 0,\dots,\frac{N}{2}$$

where $X(t)$ is the series obtained at N dates $t_1,\dots,t_N$ and $\nu_k = \frac{k}{N}$ are the Fourier frequencies.

- For a colored noise, the periodogram values at different frequencies are not independent in the case of a small N, but, they are **asymptotically** independent for large N (for $N/2$ natural frequencies).
- For a finite N, the periodogram is a biased estimator of the DSP (it means that the noise does not varie around the true DSP) but it is asymptotically non-biased.

Proposed Periodogram Calibration

To account for the frequency-dependent noise DSP, we normalize $P_r(\nu_k)$ by an averaged periodogram $P_{moy}(\nu_k)$ obtained from L independent simulated noise realizations.

$$\text{Normalized periodogram :} \qquad P_r(\nu_k) := \frac{P(\nu_k)}{P_{moy}(\nu_k)}$$

Test of planet detection

$$\text{Tested hypotheses :} \begin{cases} H_0 \ : \ X(t_j) = \epsilon(t_j) & \epsilon\text{: noise with partially unknown DSP } S(\nu_k) \\ H_1 \ : \ X(t_j) = \displaystyle\sum_{i=1}^{N_s} \alpha_i \sin(2\pi f_i + \phi_i) + \epsilon(t_j) & X\text{: the planetary signals in the same noise.} \end{cases}$$

N_s is the unknown (but presumably low) number of sinusoids with unknown amplitude α_i, frequency f_i and phase ϕ_i.

Proposed test : $T_r = \max_{\nu_k}[P_r(\nu_k)] \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$

where γ is the threshold which controls the false alarm probability (P_{FA}).
Picking the max is powerful in the case of weak amplitudes α_i and low N_s (*cf.*[4]).

Analytical results in the asymptotic regime

Null hypothesis (H_0): there is only the colored noise in the data

- For large N, the periodogram values at different frequencies become independent. In that case, the numerator of $P_r(\nu_k|H_0) \sim \chi^2_2$ and its denominator $\sim \chi^2_{2L}$. By definition (*cf.*[5], Chap. 5, theorem 5.2.6), their ratio give a F-distribution (*cf.*[6]), so we have the **statistical control of the normalized periodogram for any colored noise :**

$$P_r(\nu_k|H_0) \sim \begin{cases} \dfrac{\chi^2_2}{\chi^2_{2L}} \sim F(2, 2L) & \text{for } k = 1,\dots,\frac{N}{2}-1, \\ \dfrac{\chi^2_{1,\lambda}}{\chi^2_L} \sim F(1, L) & \text{for } k = 0 \text{ and } k = \frac{N}{2}. \end{cases} \qquad (1)$$

- The asymptotical distribution allows us to derive the **false alarm probability** thanks to the cumulative distribution function (cdf) of the F law (ϕ_F):

$$\begin{aligned} P_{FA}(T_r,\gamma) &:= \text{Prob}(\max_k P_r(\nu_k) > \gamma|\mathcal{H}_0) = 1 - \prod_{k=1}^{N/2-1} \text{Prob}\left(P_r(\nu_k) < \gamma|\mathcal{H}_0\right) \\ &= 1 - \Big(\phi_F(\gamma, 2, 2L)\Big)^{\frac{N}{2}-1} = 1 - \Big(1 - (\frac{L}{\gamma+L})^L\Big)^{\frac{N}{2}-1}. \end{aligned} \qquad (2)$$

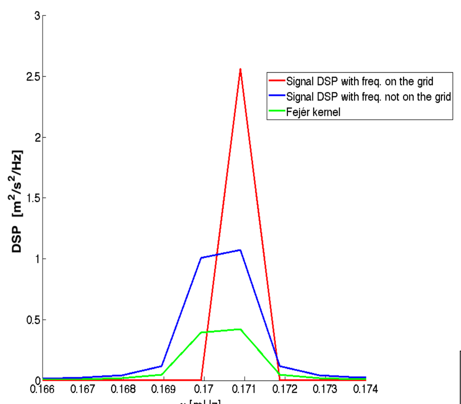

Figure 3: - Illustration of the spectral leakage. Comparaison of the DSP of a real signal not contaminated by noise $s = \alpha\sin(2\pi f_s t)$ in the case where the frequency is on the Fourier grid (red) or not (blue). The green curve represents the Fejér kernel $K_N(f_s - \nu)$.

Alternative hypothesis (H_1) : at least one planet is present

- In the same way, [7], Chap. 6, theorem 6.4 , the numerator of $P_r(\nu_k|H_1) \sim \chi_{2,\lambda}$, with $\lambda(\alpha, f, N, S(f))$ a known non-centrality parameter illustrating the fact the planet frequency does not, in general, correspond to a Fourier frequency. We obtain:

$$P(\nu_k|H_1) \sim \begin{cases} \dfrac{\chi^2_{2,\lambda}}{\chi^2_{2L}} \sim F_\lambda(2, 2L) & \text{for } k = 1,\dots,\frac{N}{2}-1, \\ \dfrac{\chi^2_{1,\lambda}}{\chi^2_L} \sim F_\lambda(1, L) & \text{for } k = 0 \text{ and } k = \frac{N}{2}. \end{cases} \qquad (3)$$

- We can deduce the **detection probability** thanks to the cdf of the F_λ law (ϕ_{F_λ}) :

$$P_{det}(T_r,\gamma) := \text{Prob}(\max_n P_r(\nu_n) > \gamma|H_1) = 1 - \prod_{k=1}^{\frac{N}{2}-1} \phi_{F_{\lambda_k}(\gamma,2,2L)} \qquad (4)$$

with :

$$\lambda_k = \frac{N}{2S(\nu_k)}\sum_{j=1}^{N_s}\alpha_j^2\Big(K_N(f_j-\nu_k)+K_N(-f_j-\nu_k)+2\frac{\sin(N\pi(f_i-\nu_k))}{N\sin(\pi(f_i-\nu_k)))}\frac{\sin(N\pi(-f_i-\nu_k))}{N\sin(\pi(-f_i-\nu_k))}\cos(2\pi(N+1)f_i+2\phi_i)\Big) \qquad (5)$$

where $K_N(\nu) = \left(\frac{\sin(N\pi\nu)}{N\sin(\pi\nu)}\right)^2$ is the spectral window which accounts for the spectral leakage of a signal with a non-Fourier frequency (see Fig.3).

Illustrations

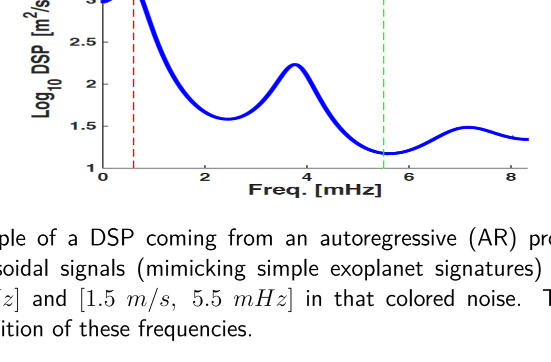

Figure 4: Example of a DSP coming from an autoregressive (AR) process. Here, we placed two sinusoidal signals (mimicking simple exoplanet signatures) characterized by $[1\ m/s,\ 0.6mHz]$ and $[1.5\ m/s,\ 5.5\ mHz]$ in that colored noise. The vertical lines illustrate the position of these frequencies.

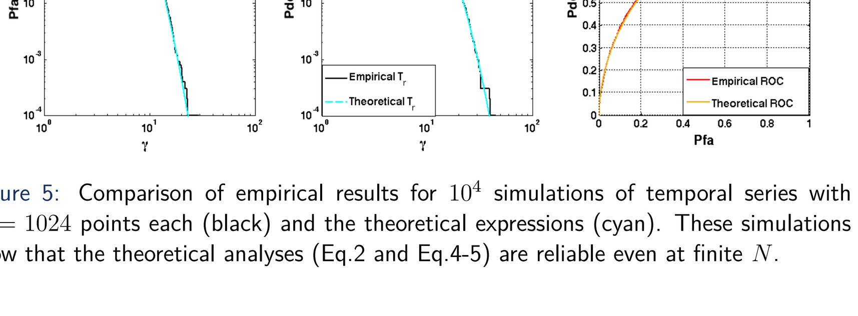

Figure 5: Comparison of empirical results for 10^4 simulations of temporal series with $N = 1024$ points each (black) and the theoretical expressions (cyan). These simulations show that the theoretical analyses (Eq.2 and Eq.4-5) are reliable even at finite N.

Fig.4 is a simulation of a colored noise DSP.
Fig.5 shows the accuracy of our theoretical results for $P_{FA}(\gamma)$(Eq.2, left panel), $P_{det}(\gamma)$ (Eq.4-5, middle) and the resulting ROC curve (i.e, P_{det} versus P_{FA} (right)). Theoretical results are checked here against Monte Carlo simulations. Although our analysis is asymptotic ($N \to \infty$), it is accurate and useful in practice for relatively low values of $N \approx 10^3$ and, of course, for larger N..
Fig.6 compares ROCs curves for classical tests and for the proposed test $T_r = \max_{\nu_k}[P_r(\nu_k)] \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$.
In the case where the signal frequency is on a noise DSP « bump », the classical tests are more powerful than others. On the contrary, when the frequency is in a « valley » of the DSP, the proposed test is better.
To understand this effect, we can investigate the frequency distribution of the test maxima under H_0 (Fig.7). **We can see that the false alarms on the « non calibrated » Fisher test** (top panel) **occur mostly in the powerful zone of the DSP contrary to the case of** T_r which shows a uniformity over all frequencies.
In conclusion, a classical test (like Fisher's) that does not take account of the stellar noise DSP are unreliable in practice. First, their detection rate vanish whenever planetary signatures do not fall close to noise DSP peaks. Second, the per-frequency false alarm rate is highly frequency dependent and the overall false alarm rate cannot be controlled. In contrast, the proposed T_r test presents a good detection power over all the whole frequency range, with performances close to the asymptotical value for L as low as $\approx 10^2$.

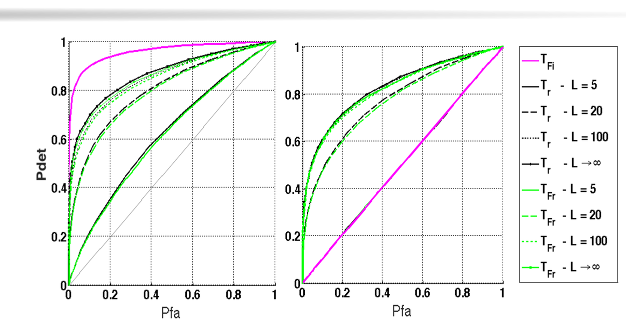

Figure 6: Comparison of ROC curves for different testing methods, resp., the Fisher test (violet, *cf*[8]), T_r (noir) and the Fisher test on P_r (green) with different L values (5, 20, 100 and ∞). The right panel is for the frequency signal in a bump on the DSP (see Fig.4) and the left panel for the frequency in a valley region of the DSP.

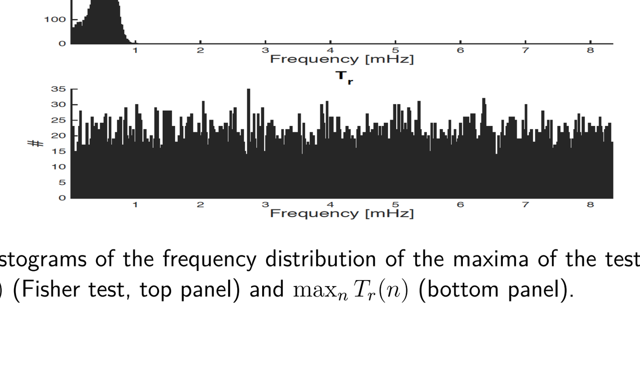

Figure 7: Histograms of the frequency distribution of the maxima of the test under H_0 : $\max_n T_{Fi}(n)$ (Fisher test, top panel) and $\max_n T_r(n)$ (bottom panel).

Conclusion

- The statistical control of the noise is necessary for a **reliable detection** of telluric planets especially when the noise is colored with partially unknown DSP !
- Here, we assumed that accurate HD simulations of the stellar noise are available and we use them to design a calibrated detection test. We use the fact that our normalized periodogram has independent values at different frequencies for large N. This allows to determine the associated false alarm and detection probability, which are independent of the stellar noise.
- The test and the corresponding performance analysis allow an accurate control of the reliability associated to a claimed detection.

Perspectives

- Improve 3D HD simulations of the stellar surface activity to effectively be able to calibrate detection tests.
- Investigate the limits of this simulation-driven approach using 3D HD simulations of a Sun-like star and periodograms from GOLF data (the Sun was observed for 10 years), in order to test the limit of the asymptotic regime.
- Generalize this asymptotical analysis to the case of irregularly sampled data.

References

[1] L. Bigot and al. *A&A*, vol. 534, no. 3, 2011.

[2] R. A. Garcia *et al. A&A*, vol. 442, pp. 385–395, 2005.

[3] A. Schuster *Journal of Geophysical Research*, vol. 3, p. 13, 1898.

[4] D. Mary and A. Ferrari in *Information Theory, 2014 IEEE International Symposium on*, pp. 561–565.

[5] D. Brillinger, *Time Series : Data Analysis and Theory, Holden Day, San Francisco, 1981.*

[6] M. Abramowitz and I. Stegun, *Spectral Analysis and Time Series, Dover Publications, 1972.*

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

[8] R. Fisher *Royal Soc. of London Proc. Series A*, vol. 125, pp. 54–59, 1929.

Acknowledgements: We acknowledge the financial support from Thales Alenia Space and the PACA region.