

## An open source scientific article

@LGRClA

### ABSTRACT

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris, consectetur id, vulputate a, magna. Donec vehicula augue eu neque, morbi tristisque senectus et netus et. Mauris ut leo, cras viverra metus rhoncus sem, nulla et lectus vestibulum. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus, morbi dolor nulla, malesuada eu, pulvinar at, mollis ac. Curabitur auctor semper nulla donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

### INTRODUCTION

Exoplanets, planets outside our solar system, are discovered at an ever-increasing rate. Beyond the study of their inner structure and atmosphere, they give a unique glimpse to extrasolar systems formation, dynamics, as well as being probes to understand exoplanets' host stars. This is particularly true for systems whose orbital plane is aligned with our line of sight, leading to observable planetary transits. Although these transit signals can be seen in the apparent flux received from their stars (light curves), they are often mixed with other astrophysical and instrumental signals, together referred to as *correlated noise*. If they can be disentangled from these nuisance signals, transits offer a powerful way to detect exoplanets. However, disentangling these signals comes with many challenges.

The simplest way to find periodic transit signals is to first clean a light curve from nuisance signals before performing the search. Indeed, widely used transit-search algorithms (namely Box-Least-Square algorithms, ) are capable of detecting transits on flat light curves only containing white noise. This strategy is widely adopted by the community, both using physically-motivated systematic models like Luger et al. (2016, 2018), or empirical filtering techniques, such as the ones described and implemented in Hippke et al. (2019). However, when systematics and variability start to resemble transits, this cleaning step (often referred to as *detrending*) strongly alters the searched signals. In such cases, the only alternative to search for transits is to perform a full-fledged modeling of the light curve, including both transits and correlated noises, and asses the likelihood of the model to the data on a wide parameter space, an approach largely avoided due to its untractable nature. Nonetheless, Kovács et al. 2016 face the question by asking: *Periodic transit and variability search with simultaneous systematics filtering: Is it worth it?*. This study discards the general benefit of using a full-fledged approach, it fails at exploring the light-curves characteristics for which it becomes necessary. While it might only represent a handful of systems, this cases are extremely valuable for the exoplanetary science community. First, variability may be associated with star-spots that can be probed with the help of planetary transits. A better understanding of these structures benefit both the study of stellar atmospheres and their concerning impact on planetary atmosphere retrievals. Second, the growing interest of the community for ultra-cool dwarf stars comes with observations featuring enhanced red noise, stellar variability and lower transit SNR. Hence, these hidden system are pristine.

In this paper we present nuance, an algorithm using linear models and gaussian processes to simultaneously search for transits while modeling correlated noises in a tractable way. In section 1, we describe the issues inherent to the two-step approach described earlier, and study the parameter space for which commonly used filtering techniques alter transit signals to the point of no detection. In section 2 we describe the tractable full-fledged approach on which nuance is based, and its implementation in an open-source Python package. In section 3, we asses the performance of nuance by computing the recovery rate of planetary transits injected into simulated light curves. In section 4, we use nuance in a case study, to detect a known planet observed by the T E S S but undetected by the TESS pipeline due to

its light curve characteristics. In section 5, we use nuance to search for transits in a list of .... Finally, we conclude this paper by providing avenues for the improvement of nuance, from its advantage to search for transits in ground-based observations, to its potential for the detection of multi-planetary systems affected by transit-times variations.

## 1. THE EFFECT OF CORRELATED NOISE DETRENDING ON TRANSIT SEARCH

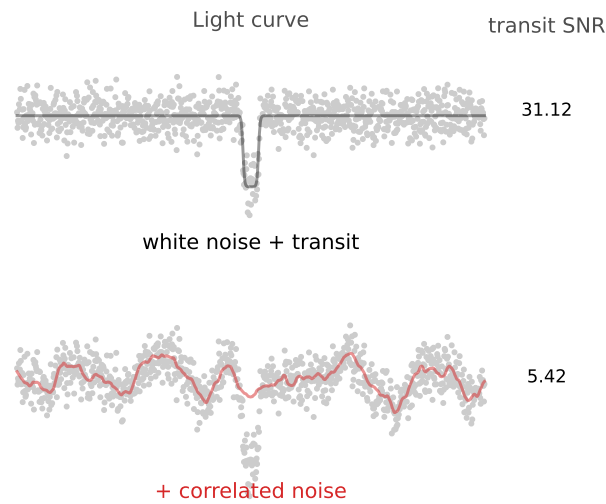
Two sources of correlated noise particularly justify the need for a detrending step before searching for transits: systematic noises, such as telescope pointing errors or detector ... and stellar variability, such as pulsations, or starspots rotations. In this section, we discuss the impact of such correlated noises on transits detectability, by studying the signal-to-noise (SNR) of a unique event, reduced to the simplified expression (Pont et al. 2006, Equation 12):

$$SNR = \frac{df}{\sqrt{\frac{\sigma_w^2}{n} + \frac{\sigma_r^2}{N_{tr}}}} \quad (1)$$

where  $df$  is the relative transit depth,  $n$  is the number of points within transit,  $N_{tr}$  the number of transits (unity here since we consider single transits), and  $\sigma_w$  and  $\sigma_c$  are the white and correlated noise standard deviations, the subscript.

### 1.1. Instrumental noises

The study from Pont et al. 2006 highlight the importance of accounting for correlated noise ( $\sigma_r$ ) when estimating the yield of transit search campaigns. To demonstrate its point, we simulate a unique transit signal (using the empirical analytical model from Protopapas et al. 2005) and compute its SNR using Equation 1, both in the presence and absence of correlated noise (Figure 1).



**Figure 1.** This is a pretty visualization of the Mandelbrot set.

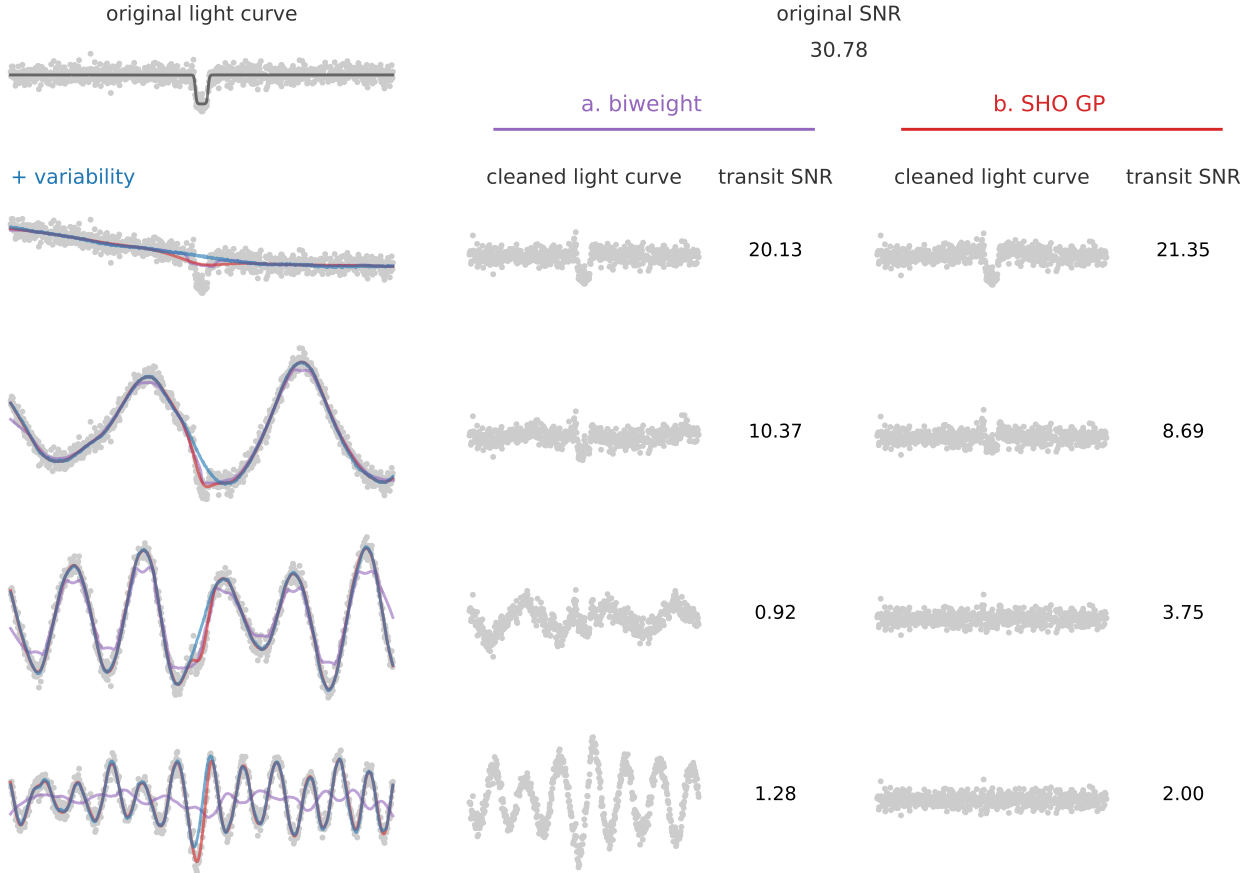
As illustrated in Figure 1, the presence of correlated noise fundamentally limits transit detectability. This issue rapidly motivated the development of systematics detrending algorithms such as the Trend Filtering Algorithm (TFA, Kovács et al. 2005, in its primary use case), SYSREM (Tamuz et al. 2005), or Pixel Level Decorrelation (PLD, Deming et al. 2015; see also EVEREST from Luger et al. 2016, 2018). Most of these methods rely on the shared nature of instrumental signals among light-curves (or neighboring pixels) such that the correction applied should not degrade the transit signal. We note that, except for TFA, these algorithms are mostly applied to space-based continuous observations, that provide continuous stellar baselines and mostly reproducible systematic signals. This is not the case for the vast majority of ground-based observations, in addition subject to periodic daytime interruptions and varying atmospheric extinction.

### 1.2. Stellar variability

Instrumental signals have the benefit to be shared among light curves of the same field, and strongly correlated with measurements from the experimental setup (like detector's temperature, pointing error, sky background or airmass).

In opposition, stellar variability is generally unknown and harder to correlate with simultaneous measurements. This gave rise to two types of treatments to reconstruct and remove stellar variability to perform a simpler transit search. One is physically-motivated and make use of Gaussian Processes (e.g. [Aigrain et al. \(2016\)](#)). The other is empirical and make use of filtering algorithms (a wide variety being described in [Hippke et al. \(2019\)](#)). In either case, the complete light curve is modeled while ignoring the presence of transits. In this section, we will study in details the effect of both approaches on transit search, by studying the degradation of a unique transit SNR for a wide variety of stellar variability characteristics.

Throughout this paper, we simulate stellar variability (and later model it) thanks to Gaussian Processes (GP), employing the physically-motivated stochastically-driven damped simple harmonic oscillator kernel (SHO) presented in [Foreman-Mackey et al. 2017](#) (see Annexe ...). We generate stellar variability thanks to the `tinygp`<sup>1</sup> Python package, providing an implementation of the scalable GP method from [Foreman-Mackey 2018](#) powered by JAX<sup>2</sup>.



**Figure 2.** This is a pretty visualization of the Mandelbrot set.

In [Figure 2](#), we simulate a light curve featuring a single transit (again using the model from [Protopapas et al. 2005](#)), in addition with stellar variability of different timescales and amplitudes. In case *a* (purple in [Figure 2](#)), we model and detrend these signals using the widely-adopted Tukey’s biweight filtering method presented in [Mosteller & Tukey 1977](#) and its implementation from `wotan`<sup>3</sup> ([Hippke et al. 2019](#)) (case *a*. in purple). In case *b* (red in [Figure 2](#)) we employ the same GP model used to simulate stellar variability to reconstruct it, and remove it from light curves ignoring

<sup>1</sup> <https://github.com/dfm/tinygp>

<sup>2</sup> <https://github.com/google/jax>

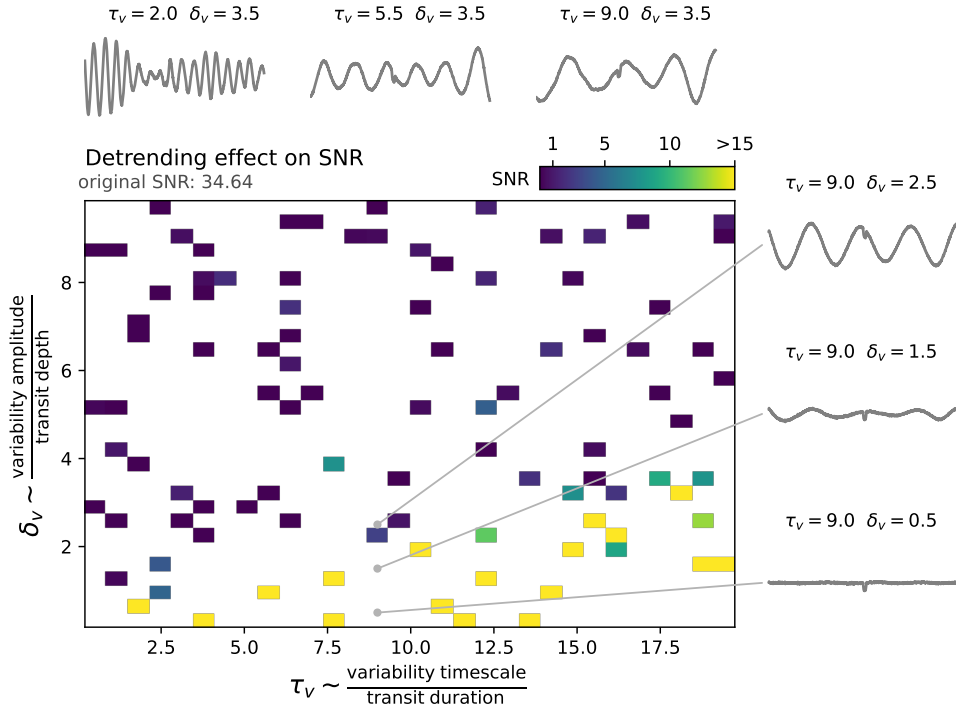
<sup>3</sup> <https://github.com/hippke/wotan>

the presence of a potential transits (as it is the case in unknown transits search). Once transits cleaned from stellar variability, we asses their remaining SNR using [Equation 1](#).

[Figure 2](#) clearly shows the effect of both detrending techniques on transits SNR, and suggests that this degradation is strongly dependant on the stellar variability characteristics encountered. In order to explore the parameter space for which detrending is the most problematic, we simulate light curves including a single transit and a much wider range of correlated noise characteristics defined with two parameters:  $\delta_v$ , the relative amplitude of the variability against transit depth; and  $\tau_v$  the relative timescale of the variability with respect to the transit duration. To follow this parametrization, we simulate the photometric variability using a GP with an SHO kernel with hyperparameters <sup>4</sup>:

$$\omega = \frac{\pi}{\tau_v \tau}, \quad \sigma = \delta_v \frac{\delta^2}{4} \quad \text{and} \quad Q = 10 \quad (2)$$

with  $\delta$  and  $\tau$  the depth and duration of the simulated transit, similar in all light curves. We fix a relatively high value for the quality factor  $Q$  in order to restrict our simulations to strongly periodic variability. The expressions of  $\omega$  and  $\sigma$  given in [Equation 2](#) correspond to a periodic signal with a period half that of the transit duration, and a peak to peak amplitude two times that of the transit depth, so that a variability with  $(\tau_v = 1, \delta_v = 1)$  strongly resembles the simulated transit signal. Then, as in [Figure 2](#), we reconstruct the variability using a biweight filter with a window length of  $3\tau$  (three times the transit duration, an optimal value according to [Hippke et al. 2019](#)). We then subtract the variability from each light-curve, estimate the resulting transit depth and compute its SNR. [Figure 3](#) shows the remaining SNR values of transits computed this way, after each variability signals with random  $(\tau_v, \delta_v)$  have been detrended.



**Figure 3.** This is a pretty visualization of the Mandelbrot set.

## 2. NUANCE

## REFERENCES

- Aigrain, S., Parviainen, H., & Pope, B. J. S. 2016, MNRAS, 459, 2408, doi: [10.1093/mnras/stw706](https://doi.org/10.1093/mnras/stw706)
- Deming, D., Knutson, H., Kammer, J., et al. 2015, ApJ, 805, 132, doi: [10.1088/0004-637X/805/2/132](https://doi.org/10.1088/0004-637X/805/2/132)
- Foreman-Mackey, D. 2018, Research Notes of the American Astronomical Society, 2, 31, doi: [10.3847/2515-5172/aaaf6c](https://doi.org/10.3847/2515-5172/aaaf6c)
- Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, AJ, 154, 220, doi: [10.3847/1538-3881/aa9332](https://doi.org/10.3847/1538-3881/aa9332)
- Hippke, M., David, T. J., Mulders, G. D., & Heller, R. 2019, AJ, 158, 143, doi: [10.3847/1538-3881/ab3984](https://doi.org/10.3847/1538-3881/ab3984)
- Kovács, G., Bakos, G., & Noyes, R. W. 2005, MNRAS, 356, 557, doi: [10.1111/j.1365-2966.2004.08479.x](https://doi.org/10.1111/j.1365-2966.2004.08479.x)
- Kovács, G., Hartman, J. D., & Bakos, G. Á. 2016, A&A, 585, A57, doi: [10.1051/0004-6361/201527124](https://doi.org/10.1051/0004-6361/201527124)
- Luger, R., Agol, E., Kruse, E., et al. 2016, AJ, 152, 100, doi: [10.3847/0004-6256/152/4/100](https://doi.org/10.3847/0004-6256/152/4/100)
- Luger, R., Kruse, E., Foreman-Mackey, D., Agol, E., & Saunders, N. 2018, AJ, 156, 99, doi: [10.3847/1538-3881/aad230](https://doi.org/10.3847/1538-3881/aad230)
- Mosteller, F., & Tukey, J. W. 1977, Data analysis and regression. A second course in statistics
- Pont, F., Zucker, S., & Queloz, D. 2006, MNRAS, 373, 231, doi: [10.1111/j.1365-2966.2006.11012.x](https://doi.org/10.1111/j.1365-2966.2006.11012.x)
- Protopapas, P., Jimenez, R., & Alcock, C. 2005, MNRAS, 362, 460, doi: [10.1111/j.1365-2966.2005.09305.x](https://doi.org/10.1111/j.1365-2966.2005.09305.x)
- Tamuz, O., Mazeh, T., & Zucker, S. 2005, MNRAS, 356, 1466, doi: [10.1111/j.1365-2966.2004.08585.x](https://doi.org/10.1111/j.1365-2966.2004.08585.x)