

Chapter 3

Detection of exoplanets around M-dwarfs

A chapter based on a paper in preparation including contributions from Daniel Foreman-Mackey and Francisco F. Pozuelos.

Although transit signals can be seen in the apparent flux received from transiting exoplanets' host stars, they are often mixed with other astrophysical and instrumental signals. If they can be disentangled from nuisance signals, transits offer a powerful way to detect exoplanets. However, disentangling these signals comes with many challenges, especially for M-dwarfs, usually observed at lower SNR and more likely to display photometric variability (e.g. Murray et al. 2020). In this dissertation, the term *correlated noise* encapsulates all nuisance signals that impact the search for transits, even if having an astrophysical origin.

Widely used transit-search algorithms (BLS types, Kovács et al. 2002) are capable of detecting transits in light curves containing only transit signals and white noise. Hence, the simplest way to find periodic transit signals is to first clean a light curve from nuisance signals before performing the search. This strategy is widely adopted by the community, both using physically-motivated systematic models like Luger et al. (2016, 2018), or filtering techniques (Jenkins et al. 2010, Hippke et al. 2019). However, when correlated noise starts resembling transits, this cleaning step (often referred to as *detrending*) is believed to degrade their detectability. In this case, the only alternative to search for transits is to perform a full-fledged modeling of the light curve, including both transits and correlated noise, and to compute the likelihood of the data to the transit model on a wide parameter space, an approach largely avoided due to its intractable nature. Nonetheless, Kovács et al. (2016) ask: *Periodic transit and variability search with simultaneous systematics filtering: Is it worth it?*. The latter study discards the benefit of using a full-fledged approach in the general case, but fails at exploring the light-curves characteristics for which it becomes necessary.

While it might only represent a handful of systems, transits hidden in correlated noise are extremely valuable for the exoplanetary science community. Indeed, correlated noise associated with stellar variability may originate from surface active regions, 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

True, but
also imp.
for det. of young planets.

This is a
little too
dismissive

(see Chapter 4 and Chapter 5).

In this chapter I present nuance, an algorithm using linear models and Gaussian processes to simultaneously search for transits while modeling correlated noise in a tractable way.

Needs
more
Context;
lit. review

Transit light curve simulations

In order to study the effect of correlated noise on transit search, this chapter relies on transit light curve simulations including a stochastically driven model of stellar variability. The following describes how such signals are modeled.

↳ realistic effects of
stellar variability and
instrumental effects.

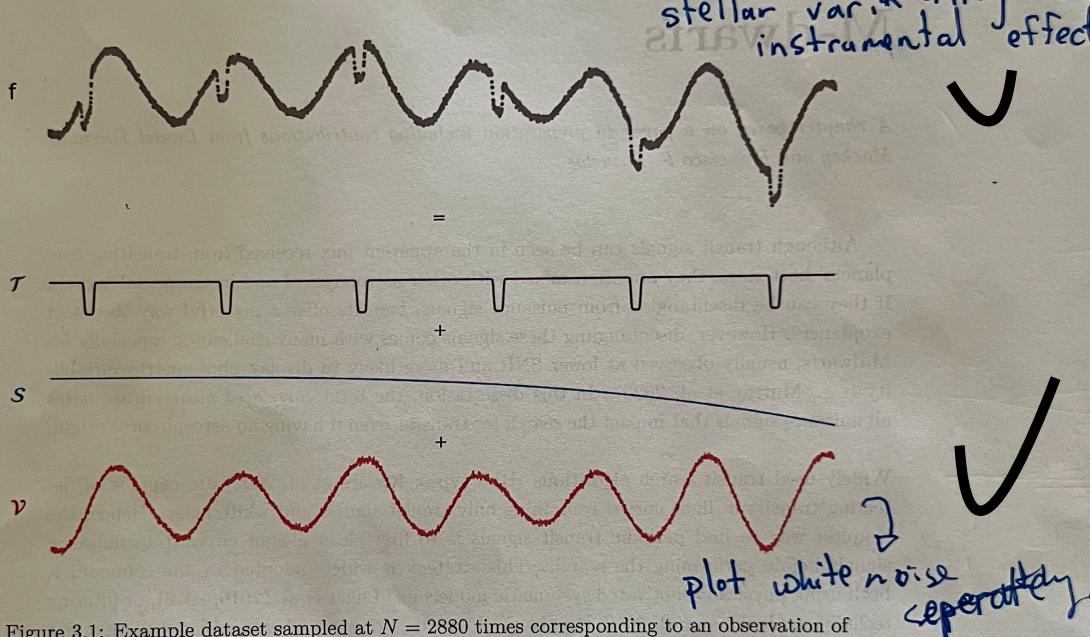


Figure 3.1: Example dataset sampled at $N = 2880$ times corresponding to an observation of 4 days with an exposure time of 2 minutes. The mean of this signal consists in a periodic transit signal of period $P = 0.7$ days, duration $D = 0.05$ days and depth of 2% (T in gray) plus instrumental signals (S in blue). Correlated noise in the form of stellar variability is simulated by modeling the covariance matrix of the signal with a Gaussian process (V in red) including a diagonal variance of 0.001^2 corresponding to white noise. This simulated signal is not intended to be physically realistic.

Let f be the simulated flux of a star sampled and arranged in the $(1 \times N)$ column-vector f associated to the column-vector of times t . f is such that

$$f \sim \mathcal{N}(\mu, C),$$

i.e. that f is drawn from a Gaussian process (GP) of mean μ and covariance matrix C (see section 1.4 for a definition of GPs). In this equation, μ is such that its i -th element is defined by $\mu_i = T(t_i) + S(t_i)$ where t_i is the i -th time of observation, T is a periodic transit

function and \mathcal{S} a function describing the instrumental part of the signal (both described below). The covariance matrix C is built such that $C_{i,j} = \nu(t_i, t_j)$ where ν is a covariance function accounting for correlated noise in the form of stellar variability with added white noise. An example of such signal is simulated and shown in Figure 3.1.

Transit signal \mathcal{T}

The periodic transit signal \mathcal{T} is simulated using the simple model described in Protopapas et al. (2005), where a transit of period P , epoch T_0 , duration D and unitary depth observed at time t is given by

$$\mathcal{T}_c(t, P, T_0, D) = \frac{1}{2} \tanh\left(c\left(\theta - \frac{1}{2}\right)\right) - \frac{1}{2} \tanh\left(c\left(\theta + \frac{1}{2}\right)\right), \quad (3.1)$$

with $\theta = \frac{P}{\pi D} \sin\left(\frac{\pi(t - T_0)}{P}\right)$,

where the dimensionless parameter c controls the roundness of the transit depth ($c \gg 1$ corresponding to a box-shaped transit as shown in Figure 3.2). This analytical model is fully empirical but easily differentiable.

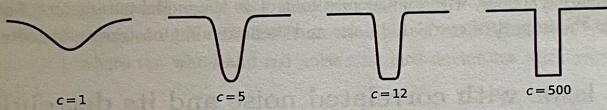


Figure 3.2: Simulations of a single transit signal (Equation 3.1) shown for different values of c .

In this chapter, all transits are simulated with $c = 12$, a value arbitrarily chosen that can be fine-tuned in real applications using the limb-darkening coefficients of a given star. The periodic transit signal \mathcal{T} seen in Figure 3.1 corresponds to $\mathcal{T} = 0.02 \times \mathcal{T}_{c=12}(t, P = 0.7, T_0 = 0.2, D = 0.05)$, all parameters in unit of days.

Instrumental signals \mathcal{S}

Instrumental signals are simulated as a linear model of M explanatory variables arranged in the $(M \times N)$ design matrix \mathbf{X} . Hence,

$$\mathcal{S} = \mathbf{w} \mathbf{X},$$

it's more typical to write the transpose, why row vectors?

where the vector \mathbf{w} contains the linear coefficients of the model. The simulated flux shown in Figure 3.1 contains a linear model where the $M = 4$ columns of the design matrix \mathbf{X} are given by $\mathbf{X}_i = t^i$ (i.e. \mathbf{X} is the Vandermonde matrix order 3 of time t) and $\mathbf{w} = [1.0 \ 0.0005 \ -0.0002 \ -0.0005]$.

Stellar variability ν

As this chapter focuses on stellar variability and its effect on transit detection, I employed a physically-motivated GP kernel, describing stellar variability through the covariance of

a stochastically-driven damped harmonic oscillator (SHO, Foreman-Mackey et al. 2017; Foreman-Mackey 2018) taking the form

$$k(\tau) = \sigma^2 \exp\left(-\frac{\omega\tau}{2Q}\right) \begin{cases} 1 + \omega\tau & \text{for } Q = 1/2 \\ \cosh(f\omega\tau/2Q) + \sinh(f\omega\tau/2Q)/f & \text{for } Q < 1/2 \\ \cos(g\omega\tau/2Q) + \sin(g\omega\tau/2Q)/g & \text{for } Q > 1/2 \end{cases} \quad (3.2)$$

where $\tau = |t_i - t_j|$, $f = \sqrt{1 - 4Q^2}$ and $g = \sqrt{4Q^2 - 1}$

needs to
be more
clear w/
what is "V"?
The kernel or
the function
values?

where Q is the quality factor of the oscillator, ω its pulsation and σ the amplitude of the kernel function. GP computations in this chapter use the implementation from `tinygp`¹, a Python package exposing the quasi-separable kernels from Foreman-Mackey (2018) and powered by JAX². The stellar variability signal in Figure 3.1 has been sampled from a GP with an SHO kernel of parameters $\omega = \pi/6D$ (i.e. a period equal to 12 times the duration D of the simulated transit), $Q = 45$ and $\sigma = 0.02$, the depth of the simulated transit. An extra term $\sigma_f^2 = 0.001^2$ is added to the diagonal of the covariance matrix, corresponding to the variance of the simulated measurement f and leading to the white noise observed in Figure 3.1.

Using these simulated signals, with an absolute control on the model parameters, I study in the next section the impact of correlated noise and its detrending on transits detectability.

3.1 The issue with correlated noise and its detrending

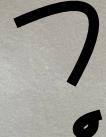
Two sources of correlated noise particularly justify the need for a detrending step before searching for transits: instrumental noises (such as telescope pointing errors) and stellar variability (induced by pulsations or starspots).

3.1.1 The effect of correlated noise on transits detectability

To study transits detectability, I focused on the signal-to-noise (SNR) of a unique event, reduced to the simplified expression (Pont et al. 2006, Equation 12):

$$SNR = \frac{\Delta}{\sqrt{\frac{\sigma_w^2}{n} + \frac{\sigma_c^2}{N_{tr}}}} \quad (3.3)$$

where Δ is the relative transit depth, n is the number of points within transit, N_{tr} the number of transits ($N_{tr} = 1$ here since we consider a single transit), and σ_w^2 and σ_c^2 are the white and correlated noise variances. To show the impact of correlated noise on transit detectability, I simulated a unique transit signal (using Equation 3.1) and computed its SNR using Equation 3.3, both in the presence and absence of correlated noise (Figure 3.3).



re-define
these
quantities

¹<https://github.com/dfm/tinygp>

²<https://github.com/google/jax>

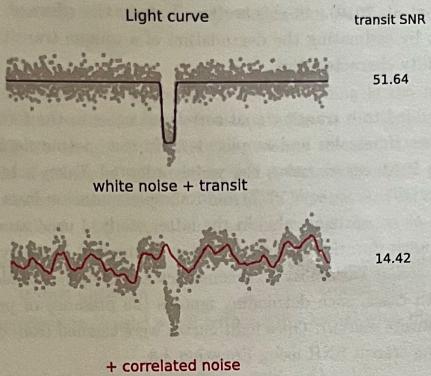


Figure 3.3: Illustration of the effect of correlated noise on a single transit SNR. I simulated a 1-hour transit of depth 1% on top of white noise (variance of 0.0015^2) as part of a 24-hours observation with an exposure time of 1 minute (top). Then, in the bottom plot, correlated noise was added to the transit signal and simulated using a GP with a Matérn-32 kernel of scale 1 hour and sigma of 0.2%. The SNR on the right of each light curve is computed using Equation 3.3, where the white and red noise standard deviations are computed as in subsection 2.2.3.

As illustrated in Figure 3.3, the presence of correlated noise strongly decreased the transit signal SNR, limiting its detection. This issue naturally motivated the development of systematic 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. Except for TFA, these algorithms were mostly applied to space-based continuous observations, that provided continuous stellar baselines and mostly reproducible systematic signals. This is not the case for the vast majority of sparse ground-based observations, in addition subject to periodic daytime interruptions and varying atmospheric extinction.

3.1.2 The effect of detrending on transits detectability

Instrumental signals often have the benefit to be shared among light curves of stars observed with the same instrument, strongly correlated with measurements from the experimental setup (like detector's temperature, pointing error, sky background or airmass time series). While not being true in general, I first assumed that the detrending of the systematic signals in light curves based on their incomplete modelling, one that ignores transit signals (because unknowns), do not affect the search for transits. In opposition, stellar variability and other astrophysical correlated noise are generally unknown and harder to correlate with simultaneous measurements. This gave rise to several treatments in order to reconstruct and detrend stellar variability. One of them is physically-motivated and makes use of Gaussian processes (e.g. Aigrain et al. 2016). Another is empirical and makes use of filtering algorithms (Jen-

ce
cite alt
not \citet

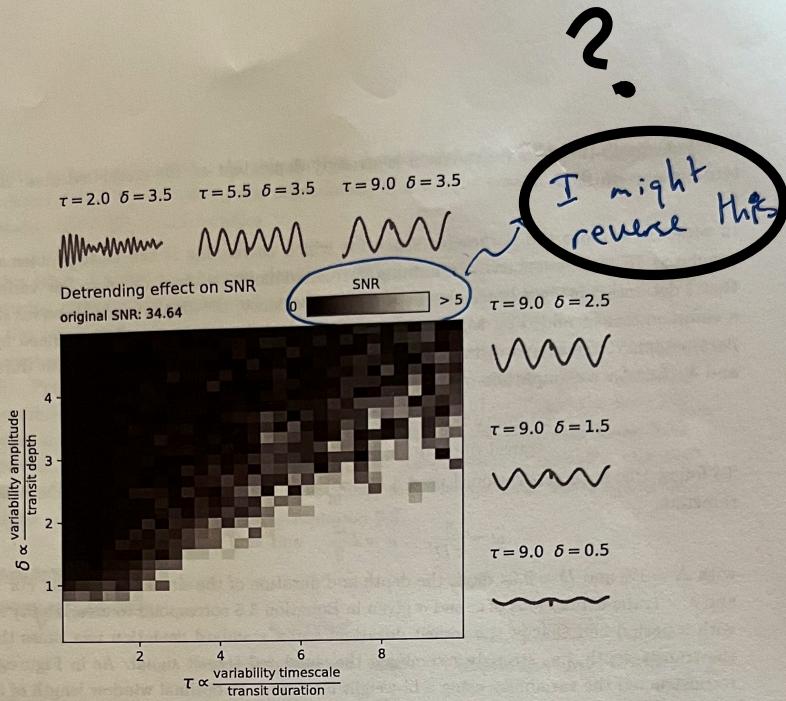


Figure 3.5: SNR of a unique transit after detrending light curves using an optimal bi-weight filter. All light curves correspond to a 2.8 days observation with a cadence of 2 minutes, and contain a unique transit of duration 1 hour and a depth of 1%, added on top of white noise with a standard deviation of 0.5% plus stellar variability. Light curves at the top and right side of the central plot are shown with their corresponding τ and δ values. The amplitude of the variability increases with δ and the timescale of the variability increases with τ .

3.2 nuance

nuance is an algorithm capable of searching for planetary transits in light curves containing correlated noise, such as instrumental signals and stellar photometric variability.

Let assume that the flux f of a star is observed and arranged in the $(1 \times N)$ column-vector f , associated to the column-vector of times t . This flux contains instrumental signals, stellar variability and a periodic transit signal that we wish to uncover (see Figure 3.1). I assume that a set of M observed measurements (such as the position of the star on the detector, the FWHM of the PSF or the sky background) taken at the same time as the flux can be treated as explanatory variables for f . These measurements are arranged in the $(M \times N)$ matrix X , later called *design matrix*.

this looks like a row vector to me...
Transpose!

Ideally, we would detect the periodic transit signal in this flux by sampling the posterior likelihood of this data to a full-fledge model including stellar variability (more generally correlated noise), instrumental systematic signals (modeled with explanatory variables), and a periodic transit signal of period P , epoch T_0 , duration D and depth Δ . We would then

reduce the posterior likelihood to $p(\mathbf{f}|P)$, its marginalized version over all parameters except the period P , producing a transit search periodogram $\mathcal{Q}(P)$. However, this approach has two issues: It is highly intractable, and it may lead to multimodal distributions that are hard to interpret.

Given a period P , we instead want to compute the likelihood of a periodic transit signal at the maximum likelihood parameters \hat{T}_0 , \hat{D} and $\hat{\Delta}$, i.e the periodogram

$$\mathcal{Q}(P) = p(\mathbf{f}|P, \hat{T}_0, \hat{D}, \hat{\Delta}) \quad (3.6)$$

This is done by adopting the strategy of Foreman-Mackey et al. (2015), and separate the transit search into two components: the linear search and the periodic search. During the linear search, the likelihood of a single non-periodic transit is computed for a grid of epochs, durations and depths. Then, the periodic search consists in combining these likelihoods to compute the likelihood of the data given a periodic transit signal for a range of periods. These combined likelihoods yield a transit-search periodogram on which the periodic transit detection is based. nuance differs from Foreman-Mackey et al. (2015) and other existing transit search algorithms as it models the covariance of the light curve with a GP, accounting for correlated noise (especially in the form of stellar variability) while keeping the model linear and tractable. This way, nuance searches for transits while, at the same time, modeling correlated noise, avoiding the detrending step that degrades transit signals SNR.

The approach employed by nuance (a two-step approach proposed by Foreman-Mackey et al. (2015)), shares similarities with the approach of Jenkins et al. (2010), where a single event statistic is computed and combined into a multiple event statistics.

3.2.1 The linear search

During the linear search, the goal is to compute the likelihood $p(\mathbf{f}|T, D, \Delta)$ of the data given a single non-periodic transit signal of epoch T , duration D and depth Δ , for a grid of epochs, durations and depths.

To account for correlated noise, the light curve f is modeled as being drawn from a GP, like the simulated light curves described in the introduction of this chapter, such that

$$\mathbf{f} \sim \mathcal{N}(\mathbf{w}\mathbf{X}, \Sigma),$$

with mean $\mathbf{w}\mathbf{X}$ (i.e. a linear model of the M explanatory variables with coefficients \mathbf{w}) and covariance Σ . To account for the presence of a single non-periodic transit of epoch T and duration D , this signal is computed and appended as the last column of the design matrix \mathbf{X} , using the simple transit model from Protopapas et al. (2005) with a unitary depth (Equation 3.1). This way, the transit signal is part of the linear model and its depth Δ can be solved linearly. Under this assumption, the log-likelihood of the data given a single non-periodic transit is (Rasmussen & Williams, 2005)

$$\ln p(\mathbf{f}|I) = -\frac{1}{2}(\mathbf{f} - \mathbf{w}\mathbf{X})^T \Sigma^{-1}(\mathbf{f} - \mathbf{w}\mathbf{X}) - \frac{1}{2} \ln |\Sigma| - \frac{N}{2} \ln 2\pi, \quad (3.7)$$

where the parameters vector \mathbf{w} and their errors σ are computed using the generalized least-square solution

$$\mathbf{w} = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \mathbf{f} \quad \text{and} \quad \sigma = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1}, \quad (3.8)$$

↑ Present
this point
(i.e. linear
model for
depth)
earlier!
when talking
about transit
model at
first.

- nuance, using its implementation from the Python package described in the previous section.
- BLS, the box-least-square algorithm (Kovács et al., 2002), using the astropy BoxLeastSquares implementation⁹.
- TLS, the transit-least-square algorithm (Hippke & Heller, 2019), an optimized version of the BLS algorithm implemented in the transitleastssquares Python package¹⁰. For this study, a modified version of the transitleastssquares package is employed, allowing to search for transits using a custom range of periods and durations.

For all of these methods, 3000 trial periods from 0.2 to 2.6 days are searched, with a single trial duration fixed to the unique known duration of 50 minutes. A transit signal is considered detected if the absolute difference between the injected and the recovered period is less than 0.01 day. To ease the detection criteria, transit periods recovered at half or twice the injected period (aka *aliases*) are considered as being detected. For this reason, detected transit epochs are not considered. Results from this *injection-recovery* procedure are shown in Figure 3.9.

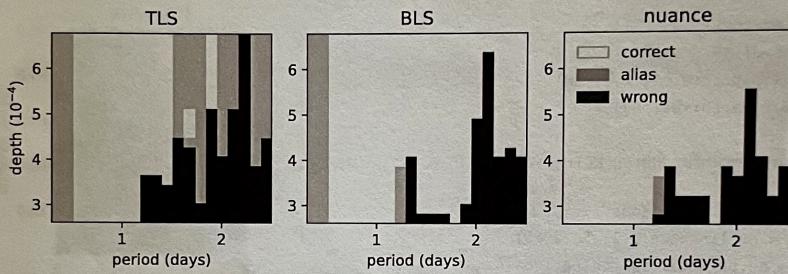


Figure 3.9: Injection-recovery of 20×20 individual transit signals in a flat light curve with only white noise (whose characteristics are described in the main text), using TLS, BLS and nuance. Transits recovered with the correct injected period correspond to the white area of the plot, while transits recovered at half or twice the injected period are shown in gray (aliases). Transit signals that were not detected are shown in black. Note that for each period and depth, a single transit epoch is used, explaining the sharp edges and discontinuities in the plot.

These results show that there is no perfect match between the detection capabilities of nuance compared to BLS and TLS. However, nuance and BLS seems to have comparable performances, compared to TLS that is less performant for lower SNR transits and more sensitive to period aliases (i.e. half or twice the injected periods). At periods lower than 0.5 days, there seem to be an advantage for nuance to detect transits with correct periods, compared to BLS and TLS detecting periods aliases.

Explaining why such differences are observed, especially between BLS and TLS is beyond the scope of this chapter, especially since BLS compares particularly well with nuance. Hence, in the following sections, nuance will be compared to methods employing BLS on light curves featuring correlated noise.

⁹<https://docs.astropy.org/en/stable/api/astropy.timeseries.BoxLeastSquares.html>

¹⁰<https://transitleastssquares.readthedocs.io>

3.4 nuance vs. biweight+BLS on simulated light curves

The subsection 3.1.2 highlighted that nuance's full-fledged modeling capabilities may not always be necessary and may only be beneficial for certain noise characteristics, relative to the searched transit parameters. As a result, this section focuses on evaluating the performance of nuance in the relative parameter space (τ, δ) described in Equation 3.5 and see when it becomes necessary.

In this section, nuance is compared to the approach that involves removing stellar variability from the light curves before performing the search on the detrended datasets. Similar to subsection 3.1.2, an optimal bi-weight filter implemented in the wotan Python package¹¹ is used with a window size three times that of the injected transit duration. The transit search is then performed using the BLS algorithm, like in section 3.3 using the astropy BoxLeast-Squares implementation¹². This transit search strategy is widely used in the community and is denoted biweight+BLS. In what follows, the transit detection criteria are the same as the ones used in section 3.3, i.e. that a transit is considered recovered if the absolute difference between the injected and recovered period is less than 0.01 days, including periods found at half or twice the injected ones (aliases).

3.4.1 Dataset

The dataset consists in 4000 light curves simulated using the model described in the introduction of this chapter. We simulate a common periodic transit added to all light curves, of period $P = 1.1$ days, epoch $T_0 = 0.2$ days, duration $D = 0.04$ days and depth $\Delta = 1\%$. Each light curve consist in a 4 days observation with an exposure time of 2 minutes, leading to $N = 2880$ data points with a normal error of 0.1%.

For a given pair of (τ, δ) , we simulate stellar variability using a GP with an SHO kernel of hyperparameters defined by Equation 3.5 (except for Q , described below), computed with respect to the injected transit parameters D and Δ . The same kernel is used for the search with nuance, an optimal choice on equal footing with the optimal $3 \times D$ window size of the bi-weight filter employed in the biweight+BLS search. 4000 pairs of (τ, δ) are generated such that

$$\tau \sim \mathcal{U}(0.1, 10), \quad \delta \sim \mathcal{U}(0.1, 25) \quad \text{and} \quad Q \sim \mathcal{U}(10, 100)$$

where $\mathcal{U}(a, b)$ denotes a uniform distribution of lower bound a and upper bound b .

3.4.2 Results

The results of this injection-recovery procedure are shown in Figure 3.10 and highlight particularly well the benefit of nuance against the biweight+BLS strategy on transits with relatively low depth compared to the stellar variability amplitude, and a relatively small duration compared to the stellar variability period. This empirical statement only concerns light curves with a given amount of white noise, and may vary depending on the length of the observing window or the number of transits. For this reason, quantifying for which values of (τ, δ) nuance outperforms biweight+BLS would only apply to this specific example. While more

¹¹<https://github.com/hippke/wotan>

¹²<https://docs.astropy.org/en/stable/api/astropy.timeseries.BoxLeastSquares.html>

?
This isn't a
fair comparison
You won't
know the
true kernel
and hyperparams
in practice.
How would
you fit
for them?