

UC Berkeley

UC Berkeley Previously Published Works

Title

Periodicity significance testing with null-signal templates: reassessment of PTF's SMBH binary candidates

Permalink

<https://escholarship.org/uc/item/2pc6056c>

Journal

Monthly Notices of the Royal Astronomical Society, 534(2)

ISSN

0035-8711

Authors

Robnik, Jakob

Bayer, Adrian E

Charisi, Maria

et al.

Publication Date

2024-09-25

DOI

10.1093/mnras/stae2220

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Periodicity significance testing with null-signal templates: reassessment of PTF’s SMBH binary candidates

Jakob Robnik ^{1,★}, Adrian E. Bayer ^{2,3}, Maria Charisi^{4,5}, Zoltán Haiman^{6,7}, Allison Lin⁶ and Uroš Seljak^{1,8}

¹Department of Physics, University of California, Berkeley, CA 94720, USA

²Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA

³Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York NY 10010, USA

⁴Department of Physics and Astronomy, Washington State University, Pullman, WA 99163, USA

⁵Institute of Astrophysics, FORTH, GR-71110, Heraklion, Greece

⁶Department of Astronomy, Columbia University, New York, NY 10027, USA

⁷Department of Physics, Columbia University, New York, NY 10027, USA

⁸Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 93720, USA

Accepted 2024 September 23. Received 2024 August 16; in original form 2024 July 24

ABSTRACT

Periodograms are widely employed for identifying periodicity in time series data, yet they often struggle to accurately quantify the statistical significance of detected periodic signals when the data complexity precludes reliable simulations. We develop a data-driven approach to address this challenge by introducing a null-signal template (NST). The NST is created by carefully randomizing the period of each cycle in the periodogram template, rendering it non-periodic. It has the same frequentist properties as a periodic signal template, and we show with simulations that the distribution of false positives is the same as with the original periodic template, regardless of the underlying data. Thus, performing a periodicity search with the NST acts as an effective simulation of the null (no-signal) hypothesis, without having to simulate the noise properties of the data. We apply the NST method to the supermassive black hole binaries (SMBHB) search in the Palomar Transient Factory (PTF), where Charisi et al. had previously proposed 33 high signal-to-noise candidates utilizing simulations to quantify their significance. Our approach reveals that these simulations do not capture the complexity of the real data. There are no statistically significant periodic signal detections above the non-periodic background. To improve the search sensitivity, we introduce a Gaussian quadrature based algorithm for the Bayes Factor with correlated noise as a test statistic. We show with simulations that this improves sensitivity to true signals by more than an order of magnitude. However, the Bayes Factor approach also results in no statistically significant detections in the PTF data.

Key words: methods: statistical – software: data analysis – quasars: supermassive black holes.

1 INTRODUCTION

Detecting periodicity in time series data X is an important task in many scientific applications, ranging from astronomy to economics. In the frequentist approach to hypothesis testing, one constructs a test statistic $q(X)$, which is designed to be large only when a periodic signal is present in the data. For example, for irregularly sampled time series, the Lomb–Scargle periodogram score is a common test statistic (Lomb 1976; Scargle 1982; VanderPlas 2018). For white noise containing a sinusoidal signal, the Lomb–Scargle score equals the likelihood ratio between the signal and no-signal hypotheses. However, for more complex noise and signals the correspondence to the likelihood ratio no longer holds, rendering it a sub-optimal test statistic.

A test statistic is calibrated by its distribution under the null hypothesis, \mathcal{H}_0 , i.e. the hypothesis that there is no signal Fisher

(1955). If the p-value = $P(q > q(X)|\mathcal{H}_0)$ is not sufficiently small, a signal detection cannot be claimed. Typically, multiple simulations of the data under the null hypothesis are generated and the p-value is estimated as the fraction of simulations for which $q > q(X)$. This approach is robust to the assumptions about the alternative hypothesis, i.e. hypothesis that the signal is present, and the choice of the test statistic: sub-optimal choices result in a lower probability of detecting the signal, but one still can be confident about the false positive probability. However, the approach is not robust to the assumptions about the null hypothesis. Realistic data sets are often too complex to be accurately simulated (LIGO Scientific Collaboration et al. 2004; Ricker et al. 2015; Jenkins et al. 2017), often leading to an underestimation of the p-value, which in turn leads to false detections (see for example Marchini et al. 2004; Stott, Stone & Allen 2004; Pont, Zucker & Queloz 2006; Price et al. 2006). The situation is even worse if one attempts to reduce the computational complexity and replace simulations with analytic estimates. For example, the analytic expressions provided in Baluev (2008, 2013) are only valid for independent identically distributed

* E-mail: jakob.robnik@gmail.com

(i.i.d.) Gaussian noise. Bayer & Seljak (2020) and Robnik & Seljak (2022) proposed to estimate the p-value from the Bayes Factor, but this is also sensitive to assumptions of the underlying model, since the Bayes Factor depends on it.

Data-driven approaches to quantify the p-value help to avoid this issue. For example, Bayer, Seljak & Robnik (2021) estimate the p-value of the highest peak in the periodogram using the secondary peaks in the periodogram. An alternative is to directly use the data to create effective null ‘simulations’. One such approach is bootstrapping (Ivezic et al. 2019a), where each permutation of the measurements is treated as a different realization of the null. However, this approach is only valid if the null is composed of independent and identically distributed data.

In this work, we propose a novel data-driven approach for quantifying the false positive probability in periodicity searches based on the modification of the periodic search template, rather than the data. The null-signal template (NST) is constructed from the periodic template by stretching and compressing the individual cycles of the signal. NST will assign low significance to the real (periodic) signal, but will have the same statistical properties as the original template on the null. Analysis with the null-signal template can therefore be used as an effective null simulation. This approach has been used in other fields. For example, gravitational waves trigger a coincident detection across multiple detectors, with a small delay, equal to the light crossing time between the detectors. Therefore, by fixing a large non-physical delay between the detectors, one can be sure to find only false positives (LIGO Scientific Collaboration et al. 2004). Their distribution is the same as the false positive distribution in the original search with coincident detection, and thus the search with the modified delay can be used as a substitute for the null hypothesis simulations. When applying the method to applications such as periodograms the challenge is to find NST that reproduces the same false positive rate as the actual search, while also being sufficiently orthogonal to the presence of true signals such that the resulting false positive rate is only weakly affected by the possible existence of true signal in the data.

We apply the method to the search for quasar periodicity, which can signify the presence of sub-parsec super-massive black hole binaries (SMBHBs). In recent years, systematic searches in large samples of quasar light curves from time-domain surveys have revealed around 150 candidates. For instance, Graham et al. (2015) identified 111 candidates in a sample of around 250 000 quasars from the Catalina Real-time Transient Survey (CRTS), while Charisi et al. (2016) identified 33 candidates among around 35 000 quasars from the Palomar Transient Factory (PTF). More recent searches in the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS), the Dark Energy Survey (DES) and the Zwicky Transient Facility (ZTF) detected a few more candidate systems (Liu et al. 2019; Chen et al. 2020, 2024). For a detailed summary of the current status of quasar periodicity searches, we refer the reader to Charisi et al. (2022) and D’Orazio & Charisi (2023). Here, we apply our new method to reassess the periodicity significance of the 33 SMBHB candidates identified in PTF.

In addition to the NST method, which is agnostic to the underlying noise properties, we also introduce some improvements to the standard quasar periodicity detection methods, which typically model quasar variability as a damped random walk (DRW) (Kelly, Bechtold & Siemiginowska 2009; Kozłowski et al. 2010; MacLeod et al. 2010). Previous searches for quasar periodicity (Graham et al. 2015; Charisi et al. 2016; Chen et al. 2020) used periodogram peaks as their test statistic and then assessed the p-values of the detected signals with DRW simulations. Given that the underlying assumption

in the periodogram is uncorrelated white noise, this makes the choice of this statistic sub-optimal, as the noise is correlated. In addition, priors may also improve the search sensitivity, by focusing on the regions of parameter space with higher prior. The posterior odds, often equivalent to the Bayes Factor, are known to be the optimal test statistic (Zhang 2017; Fowlie 2023) in the sense that they maximize the true positive rate (TPR) at a fixed false positive rate (FPR). However, the Bayes Factor with correlated Gaussian noise has been computed only for a few particularly interesting candidates (D’Orazio, Haiman & Schiminovich 2015; Vaughan et al. 2016; Zhu & Thrane 2020). This is because the evidence integral is usually computed with sampling algorithms such as nested sampling (Zhu & Thrane 2020; Witt et al. 2022), which is computationally intensive. Sometimes a Bayesian information criterion (BIC) is used as an easy to evaluate substitute for the Bayes Factor (Liu, Gezari & Miller 2018; Witt et al. 2022), but BIC does not take into account the prior information and accounts for the look-elsewhere effect overly simplistically leading to a suboptimal analysis (Weakliem 1999).

Exploiting the low dimensionality of the problem, we instead compute the Bayes Factor with a Gaussian quadrature scheme (Robnik & Seljak 2022), achieving the same accuracy as sampling methods at about three orders of magnitude lower computational cost. Specifically, a single light-curve analysis with our code takes around half a minute on a single CPU, while in Witt et al. (2022), several CPU hours were used. The speed-up enables the use of the Bayes Factor to search for periodicity in the entire quasar sample, which previously would have been prohibitive due to computational demands. This can be scaled to the even larger samples in future surveys, like the Rubin Observatory (Ivezic et al. 2019b; Xin & Haiman 2021) or the Roman telescope (Haiman et al. 2023), which will detect millions of quasars.

In Section 2, we present the methodology of the novel data-driven approach to use null-signal templates to create effective null simulations. In Section 3, we validate the method on synthetic light curves. In Section 4, we then present the improvements to the SMBHB search in PTF data and apply the NST to calibrate the results. The code with tutorials is available at <https://github.com/JakobRobnik/periodax>.

2 METHOD

Let us consider a time-series X , consisting of measurements $\{x_i\}_{i=1}^N$, with measurement errors $\{\sigma_i\}_{i=1}^N$, taken at times $\{t_i\}_{i=1}^N$, where N is the total number of observations. Let T be the time span of the data (i.e the baseline of the light curve). The measurements x_i are composed of noise n_i and signal $s(t_i)$, so that

$$x_i = s(t_i) + n_i, \quad (1)$$

where the signal

$$s(t) = A u \left(\frac{t}{P} - \phi \right) \quad (2)$$

is periodic, i.e. $u(x+k) = u(x)$ for integer k . P is the period, $\phi \in [0, 1)$ the phase, and A the amplitude of the periodic signal. The standard periodogram corresponds to a simple sinusoidal signal, $u(x) = \sin\{2\pi x\}$. Other examples include exoplanet transit searches, where $u(x)$ is a U-shaped transit (Robnik & Seljak 2021), or more complex SMBHB signals, like sawtooth periodicity (Duffell et al. 2020; Westernacher-Schneider et al. 2022) or periodically repeating lensing flares (D’Orazio & Stefano 2020).

We would like to know if a periodic signal is present in the data. The strategy is to design some test statistic $q(X)$. A popular test statistic for periodicity detection is the Lomb–Scargle periodogram

score:

$$q_{LS}(X|s) = \max_{A, P, \phi} \sum_{i,j=1}^N x_i \Sigma_{ij}^{-1} x_j - \Delta_i \Sigma_{ij}^{-1} \Delta_j \quad (3)$$

$$\Delta_i \equiv x_i - s(t_i|A, P, \phi),$$

where Σ is the noise covariance matrix, which is diagonal in the standard white Lomb–Scargle periodogram, $\Sigma_{ij} = \delta_{ij} \sigma_i^2$, and only accounts for the measurement errors. Another common test statistic is the likelihood ratio

$$q_{LR} = 2 \log \frac{\max_{z_1} p(X|\mathcal{H}_1, z_1)}{\max_{z_0} p(X|\mathcal{H}_0, z_0)}, \quad (4)$$

where z_i are the parameters of the \mathcal{H}_i hypothesis. Lomb–Scargle score is a special case of the likelihood ratio, namely when the noise is Gaussian and the null has no parameters. While the likelihood ratio takes the ratio of the likelihoods at the optimal parameters, the Bayes Factor takes the ratio of the pieces of evidence:

$$B = \frac{p(X|\mathcal{H}_1)}{p(X|\mathcal{H}_0)}, \quad (5)$$

which are the likelihoods averaged over the prior $p(X|\mathcal{H}_i) = \int p(X|z_i)p(z)dz_i$. Finally, the posterior odds additionally weigh the Bayes Factor with the prior odds, i.e. the ratio $p(\mathcal{H}_1)/p(\mathcal{H}_0)$ that favours the data sets where the signal is more probable a priori:

$$\frac{p(\mathcal{H}_1|X)}{p(\mathcal{H}_0|X)} = \frac{p(X|\mathcal{H}_1)}{p(X|\mathcal{H}_0)} \frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}. \quad (6)$$

Once a test statistic is chosen, it is evaluated on the data $q(X)$ and is compared to the distribution $P(q|\mathcal{H}_0)$ under the null hypothesis. If the observed value is very unlikely under the null, we can claim a discovery.

Here, we propose to estimate the null distribution by modifying the test statistic $q(X) \rightarrow \tilde{q}(X)$, such that the statistical properties of both statistics are the same under the null:

$$P(\tilde{q}|\mathcal{H}_0) \approx P(q|\mathcal{H}_0), \quad (C1)$$

while the modified statistic is only weakly triggered on the alternative

$$P(\tilde{q}|\mathcal{H}_1) \approx P(\tilde{q}|\mathcal{H}_0) \ll P(q|\mathcal{H}_1). \quad (C2)$$

A test statistic satisfying Conditions (C1) and (C2) can be used as an effective null simulation when applied to the real data, because it is distributed according to $P(q|\mathcal{H}_0)$, regardless of whether the data are actually from \mathcal{H}_0 or \mathcal{H}_1 . This is a purely data driven approach, and does not require any simulations to quantify the false positive rate.

We construct \tilde{q} by perturbing the signal template of the alternative hypothesis \mathcal{H}_1 , such that $\tilde{q}(X) = q(X|\tilde{s})$. For example, $\tilde{q}_{LR}(X)$ is the ratio of the likelihoods, but the alternative hypothesis uses the perturbed template. The null-signal template \tilde{s} is the same as the original template on short time-scales, i.e. the period of the signal, but differs significantly on time-scales comparable to the length of the light curve. The similarity of the templates on short time scales ensure that q and \tilde{q} trigger similarly on the null signal, satisfying condition (C1), while long-scale differences ensure that the overlap between the null-signal template and the original template is small, hence satisfying condition (C2). This method thus naturally works best when the number of signal cycles is large, but we provide an additional prescription which makes it valid with as few as two signal cycles.

There are multiple ways of modifying the template; in this work, we randomize the length of its cycles. The original template is periodic, so the n th cycle is the interval $[(n-1)P, nP]$. The null-signal template's cycle n will have length P_n , such that the n th

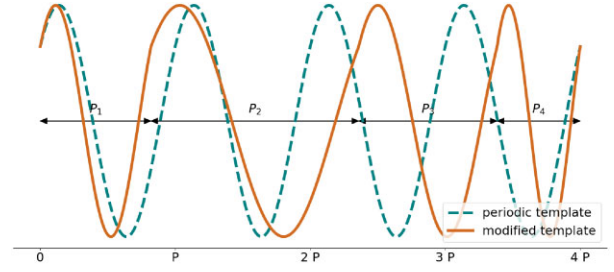


Figure 1. An example of a sinusoidal signal template (teal) and the null sinusoidal template (orange) from equation (7). The individual cycles of the null-signal template are marked with arrows.

cycle now spans the interval $[\sum_{k=1}^{n-1} P_k, \sum_{k=1}^n P_k]$. Inside each cycle, the null-signal template equals the stretched or contracted original template cycle. More formally, the null-signal template is

$$\tilde{s}(t) = A u \left(\frac{t - t_0(t)}{P_{i(t)}} - \phi \right), \quad (7)$$

where $i(t) \equiv \max\{j | \sum_{k=1}^j P_k < t\}$ and $t_0(t) \equiv \sum_{k=1}^{i(t)} P_k$. Note that, this template is continuous if u is continuous, but it is not differentiable. An example of the null-signal template is shown in Fig. 1 in orange, while the original periodic template is shown in teal.

It remains to select the lengths P_n of cycles. The original template equation (2) is a function of the signal period P , so an equivalent of the period must also be defined for the null-signal template equation (7). Furthermore, the null-signal template must vary smoothly with the period. We achieve this by requiring that the null-signal template at period P has the same number of possibly non-integer cycles as the original template with period P , which is T/P . This amounts to the following constraint

$$P_1 + P_2 + \dots + P_n + \lambda P_{n+1} = T = (n + \lambda)P. \quad (8)$$

Here, the number of cycles is decomposed into integer and fractional parts: $T/P = \lfloor T/P \rfloor + \text{frac}(T/P) \equiv n + \lambda$. The constraint from equation (8) is satisfied if we pick

$$P_n = \frac{T p_n}{\lambda p_{n+1} + \sum_{k=1}^n p_k}, \quad (9)$$

where $\{p_n\}$ can be any positive numbers. For example, one could use i.i.d. draws from the Gamma distribution which would result in periods being symmetrically Dirichlet distributed. We would like to avoid the tails of the Gamma distribution so we pick the uniform distribution, $p_n \sim \mathcal{U}(1/\sqrt{3}, \sqrt{3})$.

Having random periods can be problematic if the number of cycles is very low, because all periods can become very similar to each other by pure chance. Should this happen, the null-signal template would no longer have small overlap with the periodic template and would violate Condition (C2). To prevent this from happening, we choose p_n deterministically for low n . We define the fractional overlap between two templates as $r(s_{\text{test}}, s_{\text{true}}) = q_{LS}(s_{\text{true}}|s_{\text{test}})/q_{LS}(s_{\text{true}}|s_{\text{true}})$. This is the ratio of the Lomb–Scargle scores that the template s_{test} would assign to a signal which was actually generated by s_{true} , relative to the score that we would get using the periodic template s_{true} . Small values of the fractional overlap will ensure that Condition (C2) holds. We can achieve this by optimizing the values p_n . We do this iteratively, starting with $p_1 = 1$ without loss of generality. In the n th iteration, p_1, p_2, \dots, p_n have already been fixed. We take the time span $T = (n+1)P$ and optimize p_{n+1} such that the overlap $r(s_{\text{true}}, s_{\text{test}})$ is minimized. To make the optimized values more robust to the possible gaps in the data, we try n different gap options,

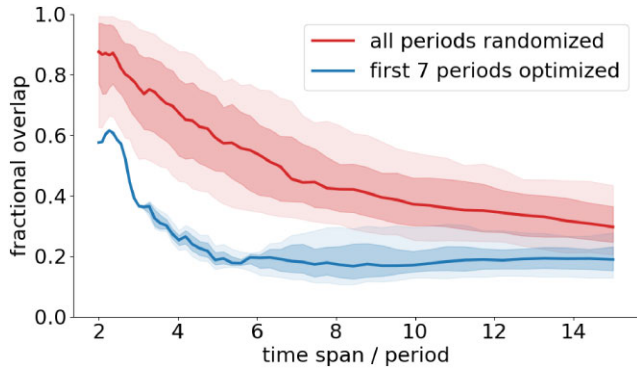


Figure 2. Lomb–Scargle periodogram score with the null-signal template (NST) relative to the score with the periodic template, i.e. fractional overlap $r(s_{\text{NST}}, s_{\text{periodic}})$. The data here are just the periodic template with no noise. Lower is better, because it ensures NST has small overlap with the periodic template and consequently obeys Condition (C2). The red line corresponds to NST if all periods are drawn from the uniform distribution. The blue line uses the fixed first seven periods that minimize the overlap and random other periods, as described in Section 2. Time sampling is 20 equally spaced points/cycle and the frequency grid for the NST is 1000 logarithmically spaced frequencies in the range $[P/10, 10P]$. The solid lines are medians over two hundred NST period randomness realizations. The stronger and the weaker band correspond to 25–75 and 10–90 quantiles, respectively. As can be seen, the optimized periods (blue) correspond to the best possible case of the random periods (red) and reduce the overlap to 60 per cent even at only two signal cycles.

each covering one signal cycle. The function that we then minimize is r , maximized over the gap choice. To compute the overlap, we use a sinusoidal template with an almost continuous time sampling (20 equally spaced points/cycle) and a dense grid for the periodic template trial frequencies (1000 logarithmically spaced frequencies in the range $[P/10, 10P]$). We optimize for the first 7 p_n and normalize them to unit average to match the scale of $p_{n>7}$. These are then fixed throughout the paper. Fig. 2 shows that this procedure yields small overlaps, even when the time span of the data is not much larger than the period of the signal.

Note that, this method only ensures that the fractional overlap between the periodic template and the NST is small, but not its absolute value. Therefore, the presence of a very large alternative hypothesis signal can compromise the first equality in condition (C2). In this case, it may be better to perform iterative analysis, such as subtracting the best-fitting signal from the data and compute the modified test statistic on residuals, otherwise the p-value estimate would be conservative. Either way this case is easy to identify, and will not be an issue for the current work, since as we will show no periodic signal candidates surpass the NST baseline.

3 VALIDATION

We now test Conditions (C1) and (C2) in a variety of synthetic examples where exact null simulations can be performed and compared to the NST effective null simulations. The examples are chosen to provide some variety of noise properties and time samplings. We consider three different noise distributions:

- (i) i.i.d. standard Gaussian,
- (ii) i.i.d. Cauchy,
- (iii) correlated Gaussian with a DRW kernel, $\tau = T/20$,

$\sigma = 1$ and measurement errors $\sigma_i = 0.2$. T is the time span of the light curve.



Figure 3. Time sampling in synthetic data sets from Section 3. Time is on the x -axis (but with different scales for different time samplings), the dots indicate times where measurements are taken. From the top to the bottom: equally spaced, equally spaced with gaps, randomly spaced. The bottom row is an example PTF time sampling for a quasar SDSS J085037.61+201337.1.

with three different time samplings:

- (i) 1000 equally spaced observations, $t_k = k$,
- (ii) equally spaced with some observations missing. We select four large regions where the data are missing and additionally independently remove each measurement with 10 per cent probability. This time sampling somewhat resembles Kepler space telescope sampling (Jenkins et al. 2017)
- (iii) 1000 uniformly random sampled $t_k \sim \mathcal{U}(0, 1)$.

Time samplings are shown in Fig. 3. For each of the 9 combinations of time samplings and noise, we do 8192 simulations. For the test statistic, we use the floating mean periodogram score from Appendix A, maximized over the trial frequencies $\{k/T\}_{k=1}^{N/2}$. The resulting cumulative distribution with the sinusoidal and the null template are shown in Figs 4 and 5. In Fig. 4, we perform noise-only simulations and show a perfect match between the two templates. Thus Condition (C1) is satisfied and is not sensitive to the noise or time sampling properties. In Fig. 5, we inject a sinusoidal signal in the data. In each realization, the signal phase $\phi \sim \mathcal{U}(0, 1)$ and frequency $1/P \sim \mathcal{U}(1/T, N/2T)$ are drawn from a uniform distribution. The amplitude of the injected signal is chosen for each combination of time sampling and noise separately, such that it significantly perturbs the distribution of the test statistic in Fig. 4. On the injected sinusoidal data, the null-signal template (teal) gives a significantly lower score to the periodic template (orange) and almost perfectly matches the pure-noise simulations with the periodic template (black). This means that Condition (C2) is also satisfied and as before is not sensitive to the noise or time sampling properties. We will further test these conditions in the next section with simulations tailored to the SMBHB search in the PTF data.

4 SMBHB SEARCH IN PTF DATA

4.1 PTF light curves

We extract quasar light curves obtained by PTF (Law et al. 2009) and apply the same cuts as in Charisi et al. (2016) to obtain a sample of 35 383 spectroscopically confirmed quasars. An example of a typical time sampling is shown in Fig. 3.

Following Charisi et al. (2016), we model the null hypothesis as a damped random walk (DRW) with additional Gaussian measurement noise and a constant offset to represent the mean magnitude. A DRW is a correlated Gaussian noise with covariance matrix

$$\Sigma_{ij}^{(DRW)} = \sigma^2 e^{-|t_i - t_j|/\tau}, \quad (10)$$

where σ is the strength of the DRW noise and τ its correlation time. We adopt a log-normal prior for both parameters,

$$\log \sigma \sim \mathcal{N}(\log 0.1, 0.2) \quad \log \tau \sim \mathcal{N}(\log 120 \text{ days}, 0.9), \quad (11)$$

following fig. 9 from Charisi et al. (2016) and MacLeod et al. (2010). The total noise is a combination of the DRW process and

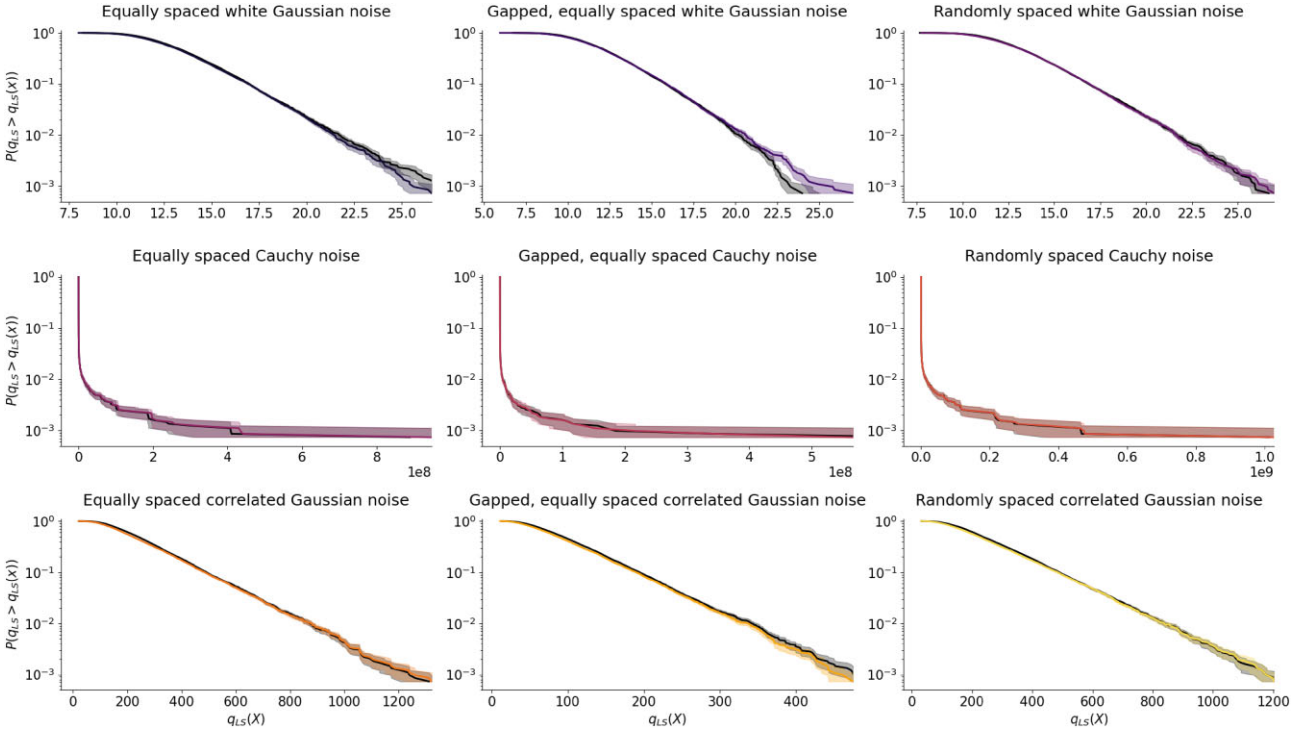


Figure 4. On noise-only simulations, the p-value with the null-signal template (coloured) agrees perfectly with the periodic template (black). A grid with all possibilities of time sampling and noise distribution from Section 3 is shown. This demonstrates that validity of Condition (C1) is not sensitive to the noise and time sampling properties. Note the large case to case differences in the signal scores. The shaded bands are statistical one σ uncertainties due to the finite number of simulations, obtained with bootstrapping.

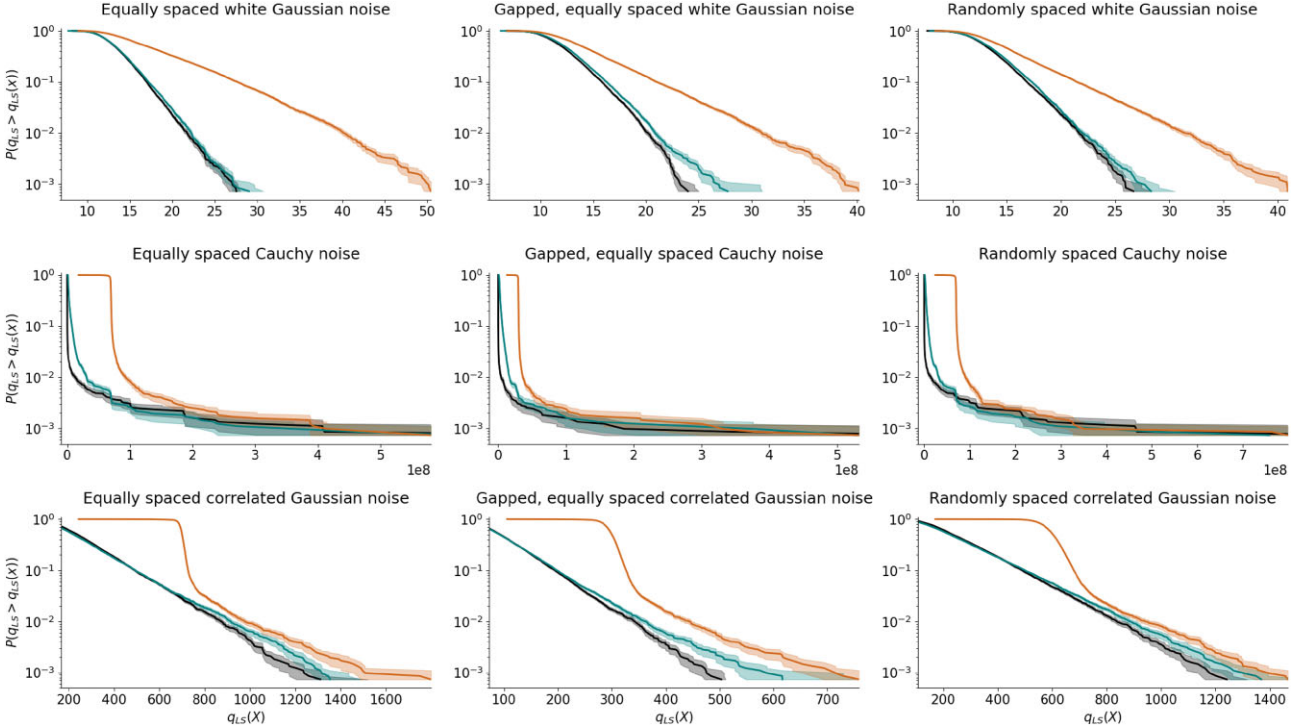


Figure 5. On noise with injected signal simulations, the score distribution with the null-signal template (teal) differs significantly from the periodic signal (orange) and matches almost perfectly the distribution with the periodic template on noise-only simulations (black). This demonstrates that validity of Condition (C2) is not sensitive to the noise and time sampling properties. A grid with all possibilities of time sampling and noise distribution from Section 3 is shown.

the measurement noise, thus the total covariance matrix is

$$\Sigma_{ij} = \Sigma_{ij}^{(DRW)} + \sigma_i^2 \delta_{ij}. \quad (12)$$

We first remove the outliers as follows: we identify the data point i as an outlier if it deviates from the Gaussian Process (GP) fit to the data by more than 3 standard deviations,

$$|x_i - \mu_{GP}(t_i)| > 3\sqrt{\sigma_i^2 + \sigma_{GP}^2(t_i)}, \quad (13)$$

where μ_{GP} and σ_{GP}^2 are the GP posterior mean and variance, respectively. We take the GP with the kernel from equation (12) with fixed $\sigma = 0.1$ and $\tau = 100$ d. We iteratively fit the GP to the light curves and identify the outliers for each new fit, repeating the procedure 5 times. We then bin together the observations taken on the same night, as in Charisi et al. (2016).

4.2 Signal prior

The alternative hypothesis (\mathcal{H}_1) is a sinusoidal signal with period P . In the quasar reference frame, the period is expected to be distributed as $p(P_{RF}) \propto P_{RF}^\alpha$, with $\alpha = 8/3$, assuming that SMBHBs are inspiraling on circular orbits and their evolution is driven by gravitational waves only (Haiman, Kocsis & Menou 2009). The observed period is related to the reference frame period by $P = P_{RF}(1+z)$, where z is the redshift of the quasar. Therefore, the observed period is also distributed as $p(P) \propto P^\alpha$. We will enforce an upper limit on the period of $P < P_{\text{cut}} = T/n_{\text{min}}$ to ensure that at least a minimal number of signal cycles $n_{\text{min}} = 2$ are observed, otherwise periodicity detection cannot be claimed. Note that, previous searches imposed a more relaxed requirement of only 1.5 cycles, which may also contribute to the detection of false positives.

The probability that a given quasar has a signal with a period in the desired range of $0 < P < P_{\text{cut}}$ is

$$P(\mathcal{H}_1) \propto \int_0^{P_{\text{cut}}/(1+z)} p(P_{RF}) dP_{RF} \propto \frac{P_{\text{cut}}^{\alpha+1}}{(1+z)^{\alpha+1}}, \quad (14)$$

where $P_{\text{cut}}/(1+z)$ is the period cutoff translated to the quasar reference frame. The prior odds $P(\mathcal{H}_1)/P(\mathcal{H}_0)$ penalizes quasars with larger redshift and shorter data span, because the required number of observed cycles limits the allowed observed period which translates to a smaller range of reference-frame periods. The proportionality constant in equation (14) further depends on quasar's brightness, because brighter quasars with more massive putative binaries chirp faster, and so spend less time at some fixed period (Peters 1964). We will ignore this effect here and assume the proportionality constant is the same for all quasars. We will fix this constant to a value that ensures $\sum_i P(\mathcal{H}_1|\text{quasar } i) = \sum_i P(\mathcal{H}_0|\text{quasar } i)$. This is in line with the standard practice in hypothesis testing, i.e. that no hypothesis is preferred a priori. The choice of the constant has no effect on the posterior odds as a test statistic, because it just shifts all scores by a constant. Finally, to make the prior smooth at the edges, we multiply $p(\log P)$ by the cosine-tapered window which has a value of 1 for $P < P_{\text{min}}/1.2$ and 0 for $P > P_{\text{min}}$. This makes it easier to maximize the posterior density during the computation of the Bayes Factor and also better aligns with our prior beliefs, which dictate a smooth prior. For the phase, we adopt the uniform prior $\phi \sim \mathcal{U}(0, 1)$. For the amplitude, we also adopt the uniform prior, but do not specify its upper bound and ignore the trials factor associated with the amplitude. This only amounts to a constant shift of the test statistic which has no relevance for the Bayes Factor as a test statistic (Robnik & Seljak 2022).

4.3 Test statistics

We try several test statistics:

- (i) Standard (uncorrelated) periodogram score, equation (3).
- (ii) Likelihood ratio, equation (4)
- (iii) Bayes Factor, equation (5)
- (iv) Posterior odds, equation (6)

We show how these can be efficiently computed in Appendices A and B. In Fig. 6, we show the receiver operating characteristic (ROC) curve for these test statistics. For this, we generated mock light curves by taking the PTF quasar data, described in Section 4.1 and replacing the magnitude measurements with realizations of DRW, whose parameters are drawn from the prior, as in equation (11). A signal with amplitude $A = 2\sigma$ and phase and period drawn from the priors in Section 4.2 is randomly injected in some light curves. The signal is injected according to the prior odds, so in around half of the light curves, but with higher probability where the prior odds are larger. To compute the ROC curve for a given test statistic, the detection threshold is varied and at each value the fraction of recovered true signals (i.e. true positive probability) is plotted against the fraction of identified false signals (i.e. false positive probability).

Fig. 6 shows that the posterior odds (yellow) are optimal in the sense that they maximize the true positive probability at a fixed false positive probability. This is the content of the Neyman–Pearson lemma for the composite hypothesis (Zhang 2017; Fowlie 2023), so we have here numerically confirmed it. Posterior odds are almost indistinguishable from the Bayes Factor (teal), meaning that the prior odds do not play a significant role in this application. The likelihood ratio (orange), which accounts for the correlated noise, but does not marginalize over the unknown parameters and does not take into account the prior distribution, closely follows the posterior odds and the Bayes Factor but is slightly less optimal. The standard periodogram score (red), which assumes white noise, is significantly suboptimal: at 0.001 false positive probability, it has a 20 times lower chance of detecting a periodic signal, because it needs a higher detection threshold to keep the number of false positives low. Most of the false positives arise from the correlated signal of the quasar light curves, which this test statistic ignores.

The take-away message is that the detection efficiency benefits significantly from incorporating the noise correlations in the test statistic. Because the Bayes Factor and posterior odds perform almost identically and the Bayes Factor is the standard choice, we will use the Bayes Factor in the remainder of the paper. Note that, one would not benefit significantly in terms of computational cost by using the likelihood-ratio instead. This is because the Gaussian quadrature integration is cheaper than the optimization, which is needed to identify the optimal parameters and is needed for both test statistics.

4.4 Null-signal template validation

Next, we repeat the test for Conditions (C1) and (C2), as in Section 3, but specifically for the SMBHB search in PTF. For this, we take the PTF light curves from Section 4.1 and replace the magnitude measurements with realizations of DRW, whose parameters are drawn from the prior in equation (11). These light curves serve as simulations of the null hypothesis \mathcal{H}_0 . As explained in 4.3, we then use the Bayes Factor as our test statistic. In Fig. 7, we show that the p-value computations with the periodic signal template (black) and the null-signal template (brown) agree perfectly, thus demonstrating Condition (C1). We also inject a sinusoidal signal in all light curves, with properties as in the previous section, i.e. with

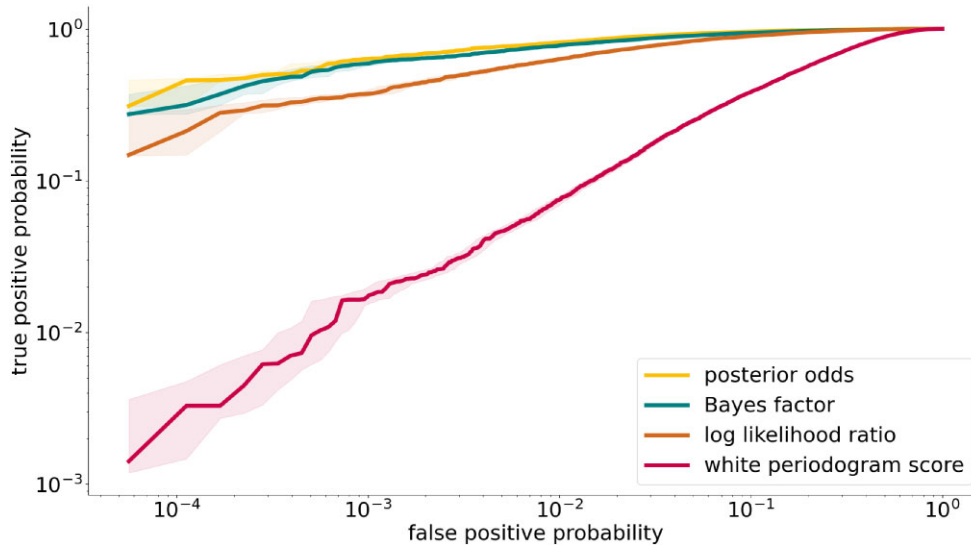


Figure 6. True positive probability as a function of false positive probability (ROC curve) for the periodicity search in the PTF data is shown for the test statistics from Section 4.3. The shaded regions are 1σ confidence bands obtained by bootstrapping. The posterior odds and the Bayes Factor perform best and are closely followed by the log-likelihood ratio. White periodogram score performs much worse.

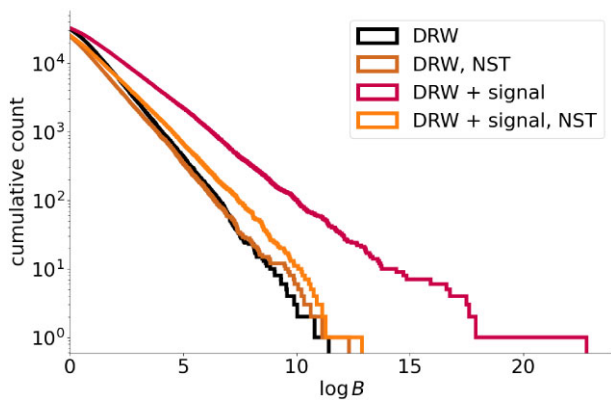


Figure 7. Cumulative number of detections as a function of the Bayes Factor test statistic on a simulation of the SMBHB search in PTF data. On noise-only simulation (DRW), the null-signal template (brown) and the periodic template (black) match perfectly, thus Condition (C1) is satisfied. On simulations of the noise with injected signal, the null template (orange) yields significantly lower scores than the periodic template (red) and matches the noise-only simulations. Thus Condition (C2) is also satisfied.

an amplitude of $A = 2\sigma$, and phase and period drawn from their prior (Section 4.2). These light curves serve as simulations of the alternative hypothesis \mathcal{H}_1 . Fig. 7 shows that the NST distribution (orange) produces significantly lower scores than the periodic signal (red), and thus Condition (C2) is satisfied.

4.5 Results

Finally, we apply all the proposed improvements for the periodicity search in the PTF light curves. In Figs 8 and 9, we show the results of this periodicity search, respectively using the Bayes Factor and white periodogram score as the test statistic. All the 33 SMBHB candidates proposed in Charisi et al. (2016) achieve $\log B < 11$, so they are no longer among the top candidates when the Bayes Factor is used as a test statistic. The top ranking candidates achieve $\log B > 20$. The two test statistics are not highly correlated, especially in the tails:

high value of the Bayes factor does not imply a high value of the white periodogram score, as shown in the right panel of Fig. 9. This is because the Bayes factor is considerably better at not triggering on the correlated noise and achieves a higher detection sensitivity. The impact on the detection results is significant.

Based solely on the raw Bayes Factor values, one might be inclined to confirm the top candidates, as even after accounting for the multiplicity of trials, $B \approx \exp 20/35383 \approx 10^4$, where we have treated all quasars equally. This is much larger than 100, a decisive evidence limit by the Jeffrey’s interpretation scale (Jeffreys 1998). Similarly, 100 DRW noise-only simulations of the entire data set analysis yield significantly lower scores than the results on the real data, see Fig. 8. DRW kernel parameters were drawn from the prior. This would suggest the p-value for these candidates is below 1 per cent.

However, both of these approaches suffer from poor modelling of the null hypothesis. This is demonstrated with the NST method, which is more robust and agnostic to the assumptions for the underlying noise. Bayes Factors and white periodogram scores, estimated with the NST, reveal that there are no statistically significant periodicity detections, compared to the non-periodic background. NST reveals much higher Bayes Factors than DRW simulations, demonstrating that the latter are a poor representation of the real variability of quasars.

In fact, NST yields higher Bayes Factor scores than the periodic template, demonstrating that there might be even more non-periodic than periodic signal data. In other words, a non-periodic signal would have been a better alternative hypothesis than a periodic signal.

In contrast, by injecting a periodic signal in the real data, we show in Fig. 10 that the periodic template would trigger a significantly higher detection than the NST. As a signal model, we take the top 30 quasars from Fig. 8. The reason why these quasars rank so highly might be some corruption of their data, so we extract the optimal signal from these quasars and inject them in 30 among the lower ranking quasars, which will here act as realistic null simulations. To ensure that the signal is injected in a similar environment, we select this second set of 30 quasars to match the original quasars in terms of the time span of the data, number of observational nights, and

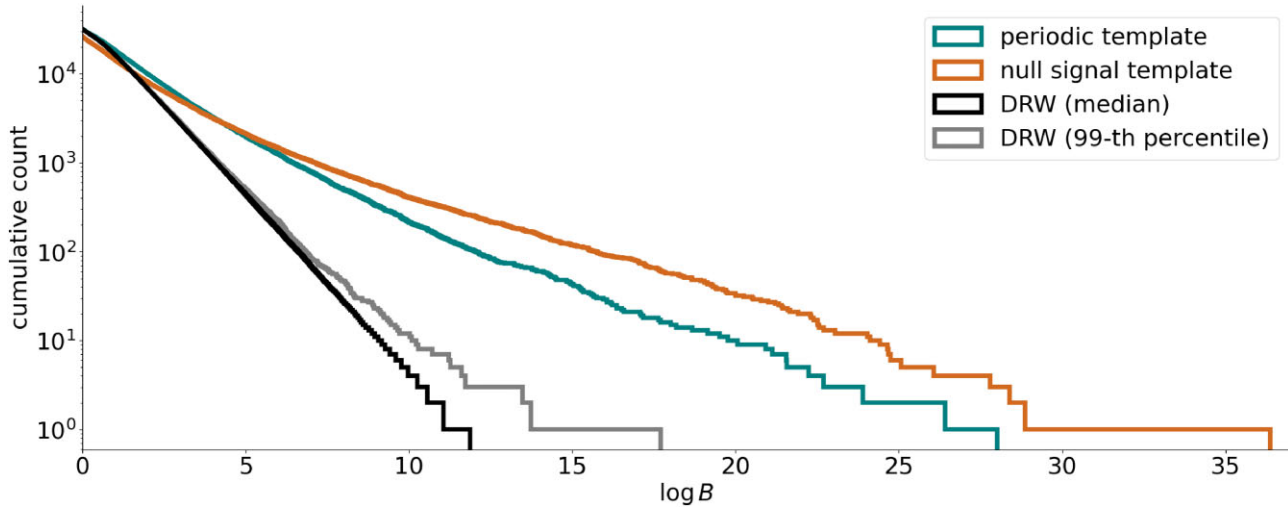


Figure 8. The results of SMBHB search in PTF data. Cumulative count of detections as a function of the detection threshold is shown. The periodic template results (teal) are to be compared with the null-signal template results (orange), which act as a null simulation. Detection does not exceed the null baseline, so a discovery cannot be claimed. In contrast, DRW simulations (black) yield significantly lower scores, which would lead to a false signal claim. We have performed 100 simulations of the entire data set analysis and sorted the candidates in each analysis by their score. The percentiles are then computed for each rank separately.

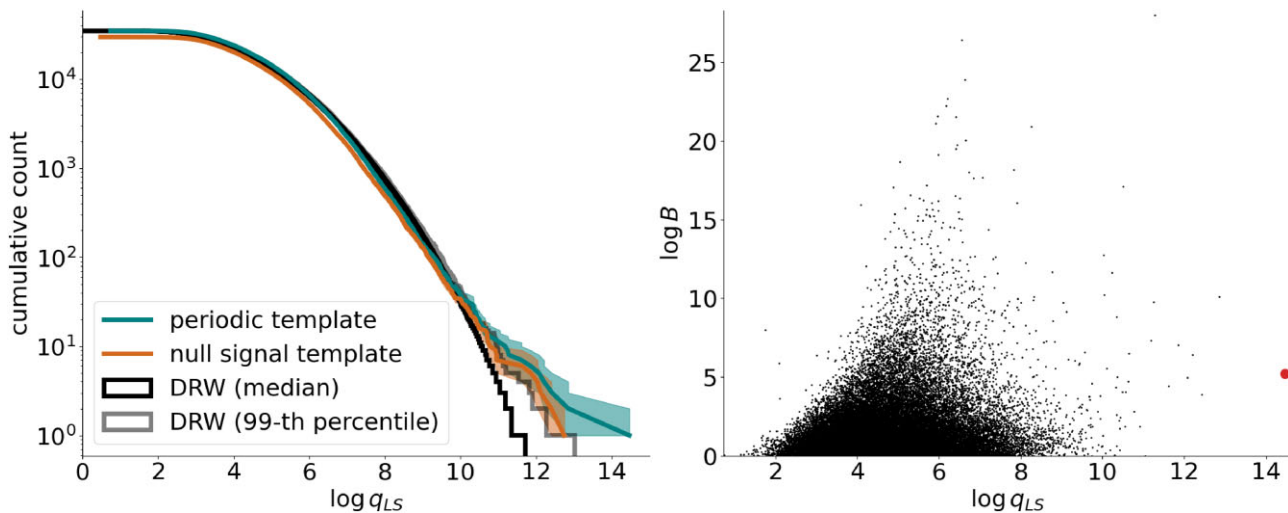


Figure 9. Left: The results of SMBHB search in PTF data with the standard white Lomb–Scargle periodogram test statistic q_{LS} from equation (3). The cumulative count of detections as a function of the detection threshold is shown. The periodic template results (teal) are to be compared with the null-signal template results (orange), which act as a null simulation. Periodic template results do not exceed the null baseline in a statistically significant manner, so a discovery cannot be claimed. The confidence regions are the quartiles obtained by bootstrapping. Here, even the (100) DRW simulations come very close to the null signal template results, showing that the majority of false positives are a result of the test statistic not incorporating the DRW noise. Right: the correlation between the white periodogram score and the Bayes factor for all the quasars, using the periodic template. Two test statistics are not highly correlated, especially in the tails: a high value of the Bayes factor does not imply high value of the white periodogram score. In particular, the candidate with the highest white periodogram score (marked in red) causes a slight deviation between the NST and periodic template results in the left panel, but has a low value of the Bayes factor. Thus, it is most likely caused by DRW not being properly modelled in the white periodogram.

parameters of the DRW kernel. We are able to match all of these parameters to less than 10 per cent deviations in 90 per cent of cases. We thus obtain 30 simulations of the alternative hypothesis which are similar to the claimed SMBHB candidates. We then analyse these data sets both with the periodic template and the NST. Fig. 10 shows that the NST gives lower scores than the periodic template, so such an SMBHB population would have been detectable if it were present in the real data.

Thus whatever the source of the NST giving a higher Bayes Factor than the periodic signal is, it cannot be caused by the presence of

real periodic signal in the data, as it would cause an opposite effect to what we found. It is however possible that the NST signal being above the periodic signal is caused by the presence of quasi-periodic signals in the data that have templates similar to the NST. However, investigating this hypothesis would require a more detailed analysis which is beyond the scope of this paper, where we only focus on periodic signals.

In appendix C, we additionally show that one can make the NST template more similar to the periodic template, such that both trigger indistinguishably on the real data, and yet the periodic

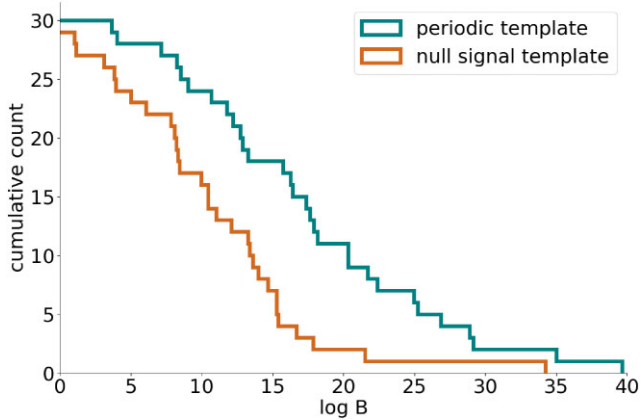


Figure 10. A demonstration that the null-signal template gives a lower significance to the periodic signal. The signal from the top 30 quasars from Fig. 8 is taken and injected in the light curves of 30 other similar quasars, as described in Section 4.5. We thus obtained a simulated data set with properties similar to the claimed SMBHB candidates. We analyse this data set both with the periodic template (teal) and the null-signal template (orange). We show that NST gives lower scores than the periodic template, so such an SMBHB population would have been detectable.

template triggers more significant detections than the NST template on periodic template injections. This test further reinforces the statement that there are no periodic detections above the non-periodic background.

In Fig. 11, we show 5 light curves with the highest Bayes Factor, both with the periodic and the null-signal template. In most cases, the light curves are similar, suggesting the periodic candidates are false positives.

5 SUMMARY

We proposed a novel method to improve the accuracy of detecting periodic signals in complex time series data by modifying the periodogram template to create a non-periodic null-signal template (NST). Since anything detected by the NST is by default a false positive, this allows us to use the real data as effective null simulations, avoiding the need for potentially inaccurate data simulations. This is

particularly useful in fields like astronomy, where the data complexity and the noise patterns are challenging to model accurately.

We validated our method through synthetic examples and applied it to the search for quasar periodicity, which have been suggested to track the presence of SMBHBs, in $\sim 35\,000$ quasar light curves from PTF. Our results showed that NST provides a robust estimate of the false positive rate, and revealed that previously proposed SMBHB candidates are likely false positives. The false positive rate in previous analyses was likely underestimated due to the assumption of DRW variability; simulations of DRW light curves cannot reproduce the NST distribution, which indicates that the DRW model may be a limited description of quasar variability, for example that its correlation time may be longer than expected, or that the distribution is non-Gaussian. Since all the systematic searches for quasar periodicity to date assume DRW model for quasar variability (Graham et al. 2015; Liu et al. 2019; Chen et al. 2020, 2024), many of the identified candidates in these searches may also be false positives, and would need to be validated with the NST method. We plan to apply our method to light curves from other surveys, like CRTS and ZTF.

Additionally, we compared several test statistics (periodogram score, Bayes Factor, posterior odds, and log likelihood ratio). We demonstrated that the Bayes Factor (and posterior odds) outperform the periodogram score, enhancing the sensitivity of periodicity searches by a factor of 20. This indicates that including the correlated noise in the test statistic is beneficial. In the future work, the test statistic could have been improved further, by computing the Bayes Factor against the true null hypothesis, which is not captured by DRW, as suggested by our results.

We also introduced a faster algorithm for computing the Bayes Factor of sinusoidal SMBHB signals against the DRW quasar noise, enabling its application to large data sets, like the upcoming LSST of the Rubin Observatory.

Looking towards future applications of NST, it is worth repeating the analysis with non-sinusoidal periodic templates, as well as with quasi-periodic templates. The challenge of these applications is again to find NSTs that satisfy conditions C1 and C2. Furthermore, while we have applied NST to a problem concerning SMBHB, the method will be useful in any application where one is searching for periodic signals, for example in exoplanet detection.

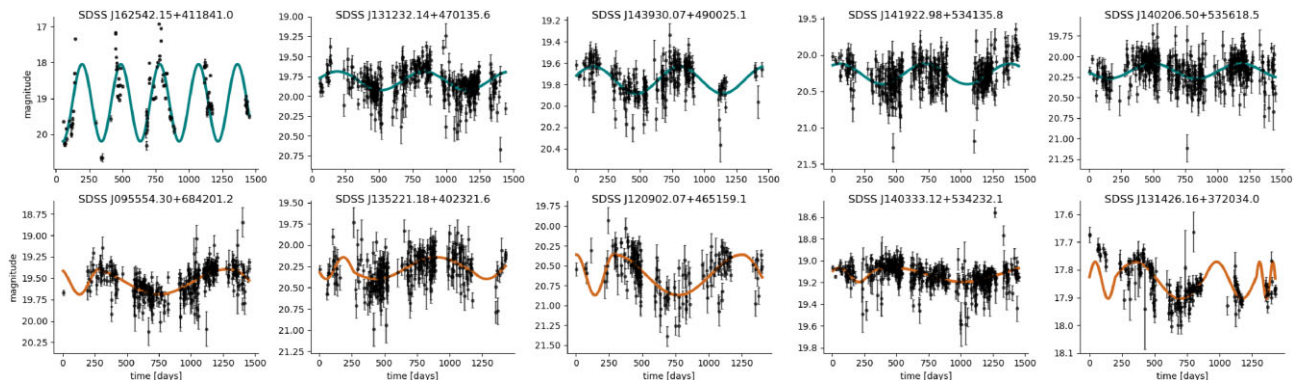


Figure 11. Top: 5 quasars with highest Bayes Factors for the SMBHB periodic signal hypothesis. Time series data and the best-fitting models are shown. Time is measured relative to MJD 54903 d. Bottom: same, but with the NST. Both periodic and NST templates visually appear a good fit to their corresponding light curves, i.e. leaving no systematic residuals. Their significance appears similar, reaffirming the statement that periodicity detection above the quasi-periodic background cannot be claimed.

ACKNOWLEDGEMENTS

This material is based upon work supported in part by the Heising-Simons Foundation grant 2021–3282 and by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under Contract No. DE-AC02-05CH11231 at Lawrence Berkeley National Laboratory to enable research for Data-intensive Machine Learning and Analysis. MC is funded by the European Union (ERC-STG-2023, MMonsters, 101117624). ZH acknowledges support by NSF grant AST-2006176 and NASA grants 80NSSC22K0822 and 80NSSC24K0440.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Baluev R. V., 2008, *MNRAS*, 385, 1279
 Baluev R. V., 2013, *MNRAS*, 429, 2052
 Bayer A. E., Seljak U., 2020, *J. Cosmol. Astropart. Phys.*, 2020, 009
 Bayer A. E., Seljak U., Robnik J., 2021, *MNRAS*, 508, 1346
 Charisi M., Bartos I., Haiman Z., Price-Whelan A. M., Graham M. J., Bellm E. C., Laher R. R., Márka S., 2016, *MNRAS*, 463, 2145
 Charisi M., Taylor S. R., Runnoe J., Bogdanovic T., Trump J. R., 2022, *MNRAS*, 510, 5929
 Chen Y.-C. et al., 2020, *MNRAS*, 499, 2245
 Chen Y.-J. et al., 2024, *MNRAS*, 527, 12154
 D’Orazio D. J., Charisi M., 2023, Observational Signatures of Supermassive Black Hole Binaries, <https://ui.adsabs.harvard.edu/abs/2023arXiv231016896D>, Access date: 28.09.2024
 D’Orazio D. J., Stefano R. D., 2020, *MNRAS*, 491, 1506
 D’Orazio D. J., Haiman Z., Schiminovich D., 2015, *Nature*, 525, 351
 Duffell P. C., D’Orazio D., Derdzinski A., Haiman Z., MacFadyen A., Rosen A. L., Zrake J., 2020, *ApJ*, 901, 25
 Fisher R., 1955, *J. R. Stat. Soc. Ser. B: Stat. Methodol.*, 17, 69
 Fowlie A., 2023, *Commun. Stat.—Theory and Methods*, 52, 5379
 Graham M. J. et al., 2015, *MNRAS*, 453, 1562
 Haiman Z., Kocsis B., Menou K., 2009, *ApJ*, 700, 1952
 Haiman Z. et al., 2023, Massive Black Hole Binaries as LISA Precursors in the Roman High Latitude Time Domain Survey, <http://arxiv.org/abs/2306.14990>, Access date: 28.09.2024
 Ivezić Z. et al., 2019a, *The Astrophysical Journal*, 873, 111
 Ivezić Z. et al., 2019b, *The Astrophysical Journal*, 873, 111
 Jeffreys H., 1998, *The Theory of Probability*. Oxford Univ. Press, Oxford
 Jenkins J. M., Tenenbaum, Seader, Burke, McCauliff, Smith, Twicken, Chandrasekaran, 2017, Kepler Science Document KSCI-19081-002, 9, Kepler Data Processing Handbook: Transiting Planet Search. <https://ui.adsabs.harvard.edu/abs/2017ksci.rept...9J/abstract>
 Karamanis M., Beutler F., Peacock J. A., Nabergoj D., Seljak U., 2022, *MNRAS*, 516, 1644
 Kelly B. C., Bechtold J., Siemiginowska A., 2009, *ApJ*, 698, 895
 Kozłowski S. et al., 2010, *ApJ*, 708, 927
 LIGO Scientific Collaboration et al., 2004, *Phys. Rev. D*, 69, 122001
 Law N. M. et al., 2009, *PASP*, 121, 1395
 Liu T., Gezari S., Miller M. C., 2018, *ApJ*, 859, L12
 Liu T. et al., 2019, *ApJ*, 884, 36
 Lomb N. R., 1976, *Ap&SS*, 39, 447
 MacLeod C. L. et al., 2010, *ApJ*, 721, 1014
 Marchini J., Cardon L. R., Phillips M. S., Donnelly P., 2004, *Nat. genetics*, 36, 512
 Peters P. C., 1964, *Phys. Rev.*, 136, B1224
 Pont F., Zucker S., Queloz D., 2006, *MNRAS*, 373, 231
 Price A. L., Patterson N. J., Plenge R. M., Weinblatt M. E., Shadick N. A., Reich D., 2006, *Nature genetics*, 38, 904
 Ricker G. R. et al., 2015, *J. Astron. Telesc. Instrum. Syst.*, 1, 014003
 Robnik J., Seljak U., 2021, *MNRAS*, 504, 5829
 Robnik J., Seljak U., 2022, *Entropy*, 24, 1328
 Scargle J. D., 1982, *ApJ*, 263, 835
 Stott P. A., Stone D. A., Allen M. R., 2004, *Nature*, 432, 610
 Stroud A. H., Secrest D., 1963, *Math. Comput.*, 17, 105
 VanderPlas J. T., 2018, *ApJS*, 236, 16
 Vaughan S., Uttley P., Markowitz A. G., Huppenkothen D., Middleton M. J., Alston W. N., Scargle J. D., Farr W. M., 2016, *MNRAS*, 461, 3145
 Weakliem D. L., 1999, *Sociol. Methods Res.*, 27, 359
 Westernacher-Schneider J. R., Zrake J., MacFadyen A., Haiman Z., 2022, *Phys. Rev. D*, 106, 103010
 Witt C. A., Charisi M., Taylor S. R., Burke-Spolaor S., 2022, *ApJ*, 936, 89
 Xin C., Haiman Z., 2021, *MNRAS*, 506, 2408
 Zhang J., 2017, *Aust. N. Z. J. Stat.*, 59, 43
 Zhu X.-J., Thrane E., 2020, *ApJ*, 900, 117

APPENDIX A: PERIODOGRAM AS A MATCHED FILTER

The likelihood of the Gaussian data under the null hypothesis is

$$-2 \log p(\mathbf{x}|\Sigma) = \mathbf{x}^T \Sigma^{-1} \mathbf{x} + \log \det 2\pi \Sigma \quad (\text{A1})$$

and

$$-2 \log p(\mathbf{x}|\mathbf{z}, \Sigma) = -2 \log p(\mathbf{x} - \mathbf{s}(\mathbf{z})|\Sigma) \quad (\text{A2})$$

under the alternative hypothesis. Σ is the noise covariance matrix which we will assume to be known in this section. \mathbf{z} are the parameters of the signal. Let the signal depend on the first M parameters linearly, we will call these parameters the amplitudes and denote them by \mathbf{a} . The signal template is then of the form

$$\mathbf{s} = \sum_{i=1}^M a^i \mathbf{s}_i(\mathbf{z}_{>M}) \equiv a^i \mathbf{s}_i(\mathbf{z}_{>M}). \quad (\text{A3})$$

As in the second step, Einstein’s convention will be used from now on: whenever there are repeated indices, the sum \sum_1^M is implied.

We are mainly be interested in the floating mean periodogram, where

$$s_1(t) = \sin 2\pi t/P \quad (\text{A4})$$

$$s_2(t) = \cos 2\pi t/P$$

$$s_3(t) = 1$$

and $z_4 = P$ is the period of the signal.

We would like to maximize the log-likelihood ratio between the two hypotheses

$$\Delta\chi^2 = 2 \log p(\mathbf{x}|\mathbf{z}, \Sigma)/p(\mathbf{x}|\Sigma) = \langle \mathbf{x}|\mathbf{x} \rangle - \langle \mathbf{x} - \mathbf{s}|\mathbf{x} - \mathbf{s} \rangle, \quad (\text{A5})$$

where we have introduced a scalar product $\langle \mathbf{x}|\mathbf{y} \rangle = \mathbf{x}^T \Sigma^{-1} \mathbf{y}$. Let’s define the metric

$$g_{ij} = \langle \mathbf{s}_i|\mathbf{s}_j \rangle \quad (\text{A6})$$

and its inverse g^{ij} , such that $g^{ij} g_{jk} = \delta_k^i$. We only consider problems with a low number of linear parameters, so inverting the metric is cheap. Let’s define the components of the data along the signal vectors as

$$x_i = \langle \mathbf{x}|\mathbf{s}_i \rangle. \quad (\text{A7})$$

With this notation the score is

$$\Delta\chi^2 = 2a^i x_i - a^i a^j g_{ij}. \quad (\text{A8})$$

Maximization over the linear parameters can be performed analytically:

$$\left. \frac{\partial \Delta \chi^2}{\partial a^i} \right|_{\tilde{a}} = 2x_i - 2g_{ij}\tilde{a}^j = 0, \quad (\text{A9})$$

which gives the optimal amplitudes

$$\tilde{a}^i = g^{ij}x_j. \quad (\text{A10})$$

The score $\Delta \chi^2$ at the optimal parameters is

$$\Delta \chi^2 = 2g^{ij}x_ix_j - g^{ij}g^{kl}x_jx_kg_{ik} = g^{ij}x_ix_j. \quad (\text{A11})$$

This is the matched filter: the data are matched against the templates s_i and inverse weighted with the noise power, here included in the definition of the scalar product. Most of the searches to date ignore the inverse covariance matrix weighting, resulting in a sub-optimal matched filter analysis.

The strategy for finding the optimal parameters \mathbf{z} is to compute the score from equation (A11) on a grid of non-linear parameters and identify the highest score.

The main cost of equation (A11) is in computing the scalar products $\langle \cdot | \cdot \rangle$. If noise is stationary with measurements taken on a regular grid, the computation can be simplified significantly, but we do not make these assumptions here. The scalar product is computed as $\langle \mathbf{u} | \mathbf{v} \rangle = \mathbf{u} \cdot \tilde{\mathbf{v}}$, where \cdot is the standard scalar product and $\tilde{\mathbf{v}}$ is the solution of the linear system $\Sigma \tilde{\mathbf{v}} = \mathbf{v}$. Since the covariance matrix is positive definite, we can decompose it using a Cholesky decomposition as $\Sigma = LL^T$, where L is lower triangular. Once the Cholesky decomposition is computed, solving the linear system is easy: first one solves $L\mathbf{v}' = \mathbf{v}$ for \mathbf{v}' and then $L^T\tilde{\mathbf{v}} = \mathbf{v}'$ for $\tilde{\mathbf{v}}$.

APPENDIX B: EFFICIENT BAYES FACTOR COMPUTATION

The difficulty in computing the Bayes Factor (equation 5) is in doing the evidence integrals

$$p(X|\mathcal{H}_i) = \int p(X|\mathbf{z}_i)p(\mathbf{z})d\mathbf{z}_i. \quad (\text{B1})$$

Linear parameters have a Gaussian posterior so can be integrated out of equation (B1) analytically. They contribute a constant that depends only on the prior volume of the linear parameters and will therefore be irrelevant if the Bayes Factor is not taken at a face value, but as a test statistic (Robnik & Seljak 2022). What remains is therefore to integrate over the non-linear parameters at optimal amplitude parameters $\mathbf{a}(\mathbf{x}, \mathbf{z}_{>M})$:

$$\tilde{p}(x|\mathcal{H}_i) = \int p(x|\hat{\mathbf{a}}(\mathbf{x}, \mathbf{z}_{>M}), \mathbf{z}_{>M})p(\mathbf{z}_{>M})d\mathbf{z}_{>M}. \quad (\text{B2})$$

Because the number of non-linear parameters is low, we can avoid the cost of MCMC by using the Gaussian quadrature scheme. Our

strategy is to find the parameters which maximize the posterior density and compute the Hessian of the negative log posterior at those parameters. This is the Laplace approximation to the posterior. It is used to set out the Gaussian quadrature scheme and integrate the posterior (Robnik & Seljak 2022). In Section 4, the null has two non-linear parameters and the alternative has one additional non-linear parameter. We will use quadrature schemes of order 6 and 7 from Stroud & Secrest (1963). We have compared these computations with pocoMC (Karamanis et al. 2022), which is a state of the art MCMC method. Both methods agree to within the MCMC error but quadrature uses 12 log-posterior evaluations in two dimensions and 27 evaluations in three dimensions, while pocoMC uses 2×10^5 in its default setting and still several thousand calls if specialized to the simple low dimensional targets.

Thus, the procedure for computing the Bayes Factor is

- (i) Maximize the null log-posterior to find the optimal parameters of the null and the Hessian at the peak.
- (ii) Using the Gaussian quadrature scheme, compute the evidence for the null.
- (iii) Take the optimal null parameters in the matched filter periodogram from Section A and find the optimal parameters of the alternative hypothesis.
- (iv) Maximize the alternative hypothesis log-posterior over all parameters of the alternative hypothesis (including the null parameters).
- (v) Using the Gaussian quadrature scheme, compute the evidence for the alternative.

APPENDIX C: INTERPOLATION FROM THE NULL TO THE PERIODIC TEMPLATE

In Fig. 8, we saw that the NST yields higher scores than the periodic template. This suggests that there is more non-periodic than periodic features in the data. As an additional test, we here make the NST perturbation from periodicity less drastic. We introduce a family of null-signal templates, parametrized by the mixture parameter β . NSTs in this family are of the form described in Section 2 and have

$$p_n = \beta p_n^{(\text{optimal})} + (1 - \beta), \quad (\text{C1})$$

where $p_n^{(\text{optimal})}$ are the values from Section 2. Thus, this family smoothly interpolates between the periodic template ($\beta = 0$) and the NST from Section 2 ($\beta = 1$). In Fig. C1, we reanalyse the real data (equivalent of Fig. 8 in the main text) and real data with injected signal (equivalent of Fig. 10 in the main text) with NST that has $\beta = 0.2$. The real data analysis shows that the NST gives the same test statistic distribution as the periodic template. The injected data demonstrates that if the periodic signal is present in the data, NST yields lower scores.

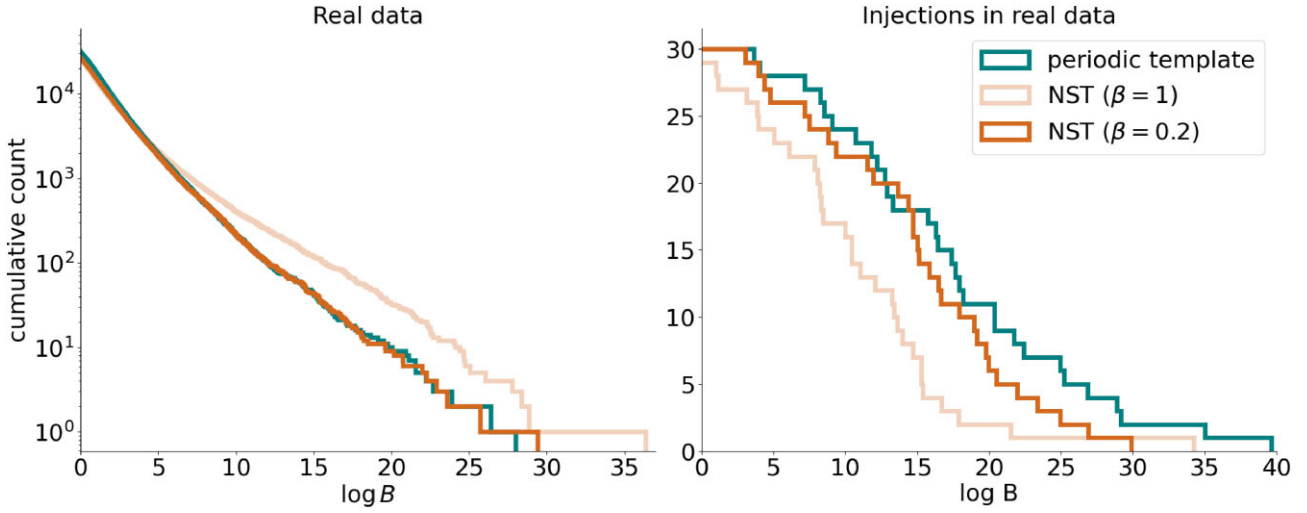


Figure C1. Detection count as a function of the Bayes Factor detection threshold. Left: periodic template and the null-signal template are compared on real data, as in Fig. 8, but with the addition of the $\beta = 0.2$ NST. $\beta = 0.2$ template yields a distribution which is practically indistinguishable from the distribution with the periodic template. Right: same but with periodic signal injection in the real data, equivalent to Fig. 10. Even at $\beta = 0.2$, NST yields noticeably lower scores than the periodic template, so such a population of periodic signals would have been detectable.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.