Toward the Unambiguous Identification of Supermassive Binary Black Holes through Bayesian Inference

Xing-Jiang Zhu1,2 and Eric Thrane1,2

1 School of Physics and Astronomy, Monash University, Clayton, VIC 3800, Australia; zhu.xingjiang@gmail.com
2 OzGrav: Australian Research Council Centre of Excellence for Gravitational Wave Discovery, Clayton, VIC 3800, Australia

Abstract

Supermassive binary black holes at subparsec orbital separations have yet to be discovered, with the possible exception of blazar OJ 287. In parallel to the global hunt for nanohertz gravitational waves from supermassive binaries using pulsar timing arrays, there has been a growing sample of candidates reported from electromagnetic surveys, particularly searches for periodic variations in the optical light curves of quasars. However, the periodicity search is prone to false positives from quasar red noise and quasiperiodic oscillations from the accretion disk of a single supermassive black hole, especially when the data span fewer than a few signal cycles. We present a Bayesian method for the detection of quasar (quasi)periodicity in the presence of red noise. We apply this method to the binary candidate PG 1302–102 and show that (a) there is very strong support (Bayes factor $>10^6$) for quasiperiodicity and (b) the data slightly favor a quasiperiodic oscillation over a sinusoidal signal, which we interpret as modest evidence against the binary black hole hypothesis. We also find that the prevalent damped random walk red-noise model is disfavored with more than 99.9% credibility. Finally, we outline future work that may enable the unambiguous identification of supermassive binary black holes.

Unified Astronomy Thesaurus concepts: Supermassive black holes (1663); Quasars (1319); Galaxy mergers (608); Bayesian statistics (1900)

1. Introduction

Supermassive binary black holes are thought to be common in the universe as natural products of galaxy mergers (Begelman et al. 1980). They are likely to be the primary sources of nanohertz gravitational waves, which have been actively searched for using pulsar timing arrays in the past decade (see, e.g., Hobbs et al. 2010; Verbiest et al. 2016; Perera et al. 2019, and references therein). Current pulsar timing arrays are sensitive to individual binaries, with component black hole masses $\gtrsim 10^9$ $M_\odot$ and binary orbital periods $\lesssim 10$ yr (or orbital separations $\lesssim 0.01$ pc), up to distances of $\sim 200$ Mpc (Zhu et al. 2014; Babak et al. 2016; Aggarwal et al. 2019); however, see Rosado et al. (2016) for the possibility of detecting high-redshift binaries. No detection has been made so far, although it has been suggested that the detection of a stochastic gravitational-wave background formed by the combined emission from binaries across the universe is likely in a few years (e.g., Taylor et al. 2016; Cordes & McLaughlin 2019).

The electromagnetic identification of subparsec supermassive binary black holes has proven to be challenging. Direct imaging of such close binaries is only possible for sources within $\sim 100$ Mpc via very long baseline radio interferometry observations. The closest binary with confident direct images, found in the radio galaxy 0402+379, has a separation of 7.3 pc (Rodriguez et al. 2006), corresponding to an orbital period of $\sim 10^4$ yr (Bansal et al. 2017); see also Kharb et al. (2017) for a 0.35 pc binary candidate at 116 Mpc in the Seyfert galaxy NGC 7674. How binaries like these can reach milliparsec orbital separations, where the emission of gravitational waves can efficiently drive the binary to merge, is referred to as the “final-parsec problem,” which remains an active area of theoretical investigation (e.g., Milosavljević & Merritt 2001; Yu 2002; Colpi 2014; Khan et al. 2016; Taylor et al. 2017; Goicovic et al. 2018; Ryu et al. 2018; Chen et al. 2019; Muñoz et al. 2020). Finding a subparsec supermassive binary black hole would not only provide insight into the final-parsec problem but also shed light on the expected stochastic gravitational-wave signal strength for pulsar timing arrays (Goulding et al. 2019; Zhu et al. 2019).

Periodically variable active galactic nuclei provide an interesting class of candidates for subparsec supermassive binary black holes. The most prominent object is OJ 287, a blazar that exhibits 12 yr quasiperiodic outbursts in optical light curves that date back to the 19th century (Sillanpaa et al. 1988; Valtonen et al. 2008). This periodicity has been interpreted as the secondary black hole crossing the accretion disk of the primary in an eccentric orbit (Valtonen et al. 2016; Dey et al. 2019). In the past 5 yr, with the growth of time-domain astronomy, there have been a large number of subparsec supermassive binary black hole candidates claimed by various groups.

Based on the Catalina Real-time Transient Survey (CRTS), Graham et al. (2015a) put forward 111 binary black hole candidates from an optical variability analysis of 243,500 quasars. Among them, PG 1302–102 was the most significant candidate, with a measured period of 1884 days (Graham et al. 2015b). This periodicity has been attributed to the relativistic Doppler boosting of emission from a minidisk around the secondary black hole (D’Orazio et al. 2015). Charisi et al. (2016) reported 33 binary candidates from a periodicity search in a sample of 35,383 quasars in the photometric database of the Palomar Transient Factory. Liu et al. (2015) identified a periodic variability of 542 days in

8 After this paper was submitted, Laine et al. (2020) reported Spitzer observations of the predicted Eddington flare from OJ 287, adding significant weight to the theory that it is, in fact, a supermassive binary black hole.
quasar PSO J334.2+01.4 using data from the Pan-STARRS1 Medium Deep Survey. However, such a detection was found to be insignificant in a subsequent analysis of extended data (Liu et al. 2016); see Liu et al. (2019) for a systematic search over 9000 quasars that resulted in one candidate in their extensive analysis.

Following the periodicity report of PG 1302−102, Vaughan et al. (2016) cautioned that the stochastic variability of normal quasars (i.e., those that do not host binary black holes) can resemble a periodic feature in light curves that span only a few periods and highlighted the importance of careful evaluation of the false-alarm rate. Through Bayesian model comparison, Vaughan et al. (2016) found that a red-noise model is significantly favored over a sinusoidal model for PG 1302−102 based on the CRTS data. More recently, Liu et al. (2018) revisited the periodicity of PG 1302−102 using additional data from the All-Sky Automated Survey for Supernovae (ASAS-SN). They employed a maximum-likelihood method to search for a periodic signal in the presence of quasar red noise and found that the inclusion of ASAS-SN data reduced the periodicity significance. Therefore, Liu et al. (2018) concluded that the binary black hole model was disfavored for PG 1302−102. Kovačević et al. (2019) proposed a model that posits a cold spot in the accretion disk of the primary black hole. Such a model can produce a perturbed sinusoidal feature and thus explain the apparent decrease in periodicity significance found by Liu et al. (2018).

Here we propose a fully Bayesian framework for the identification of supermassive binary black hole candidates in time-domain electromagnetic surveys. It is capable of dealing with generic signal forms in the presence of red noise. Our work improves on previous studies in several ways. First, our method allows the inference of noise properties and signal parameters simultaneously and thus accounts for potential covariance between a periodic signal and quasar red noise. Second, it is robust to offsets in data collected with different surveys and possible over/underestimation of measurement uncertainties. Third, we adopt a general form of quasar red noise and search for deviation from the commonly assumed damped random walk (DRW) model (Kelly et al. 2009). We also compare the sinusoidal signal hypothesis against a quasiperiodic oscillation (QPO) model based on behavior observed in many stellar-mass black hole X-ray binaries (see, e.g., Motta 2016, for a recent review).

The remainder of this paper is organized as follows. In Section 2, we describe the Bayesian inference framework and introduce our signal and noise models. In Section 3, we apply the method to PG 1302−102 with data from CRTS, ASAS-SN, and the Lincoln Near-Earth Asteroid Research (LINEAR) survey and present our analysis results. In Section 4 we discuss various aspects of a periodicity search through simulations. Last, we provide concluding remarks and outline directions for future work in Section 5.

2. Bayesian Inference and Model Selection of Time Series

In this section, we describe the framework of Bayesian inference and model selection for the analysis of time-series data. We focus on the case of searching for periodicity in quasar light curves—quasar brightness measurements as a function of time. Assuming stationary Gaussian noise, the likelihood function for a quasar light curve is

\[
\mathcal{L}(d|\theta_n, \theta_s, m) = \frac{1}{\sqrt{(2\pi)^N|\mathbf{C}|}} \exp \left[ -\frac{1}{2}(d - m - s)^T \mathbf{C}^{-1}(d - m - s) \right].
\]

where \(d\) is the time-series light-curve data with length \(N\), and \(\theta_n\) and \(\theta_s\) include the noise and signal parameters, respectively. In this work, we use measurements of optical magnitudes (i.e., the logarithmic light curve). The data \(d\) are modeled as

\[
d = n + m + s,
\]

where \(n\) is the noise vector, which contains measurement uncertainties and additional intrinsic stochastic quasar variability; \(m\) is a constant vector with identical entries of \(m\), where \(m\) accounts for the mean magnitude and any constant offset (e.g., constant level of contamination due to host galaxy light); and the signal vector \(s\) is given by

\[
s(t) = A \sin(2\pi f_0 t + \phi).
\]

The signal parameters are \(\theta_s = \{A, \phi, f_0\}\), where the signal frequency \(f_0\) is related to the period by \(f_0 = 1/T_0\). In Equation (1), \(C_{ij} = \langle n_in_j \rangle\) is the noise covariance matrix, which contains two components, \(C = C^\nu + C^\epsilon\), where \(C^\nu\) is a diagonal matrix that represents white noise, and \(C^\epsilon\) accounts for the stochastic quasar variability, usually termed “red noise.”

The white-noise matrix takes the form

\[
C^\nu_{ij} = (\nu \sigma_i)^2 \delta_{ij},
\]

where \(\sigma_i\) is the measurement uncertainty for the \(i\)th observation, \(\nu\) is a scale factor used to quantify the over/underestimation of measurement uncertainties,\(^4\) and \(\delta_{ij}\) is the Kronecker delta function.

In the DRW model (Kelly et al. 2009), the covariance function is an exponential function. The red-noise covariance matrix is

\[
C^\epsilon_{ij} = \frac{1}{2} \hat{\sigma}^2 \tau_0 \exp \left( -\frac{\tau_{ij}}{\tau_0} \right).
\]

where \(\hat{\sigma}^2\) is the intrinsic variance between observations on short timescales (~1 day), \(\tau_0\) is the damping timescale, and \(\tau_{ij} \equiv |t_i - t_j|\). The noise power spectral density of the DRW model is

\[
P(f) = \frac{2\hat{\sigma}^2 \tau_0^2}{1 + (2\pi\tau_0 f)^2}.
\]

To account for the possibility that the quasar red noise deviates from the DRW model (e.g., Zu et al. 2013; Guo et al. 2017), we also consider the following form of the noise

\(^4\) This is equivalent to the Error FACtor (EFAC) parameter used in pulsar timing data analysis. An additional Error added in QUADrature (EQUAD) parameter could be introduced to account for potential systematic errors that are not included in measurement uncertainties.
covariance matrix:

\[ C_{ij}^r = \frac{1}{2} \sigma_i^2 \sigma_j^2 \exp \left[ -\left( \frac{T_{ij}}{\tau_0} \right)^2 \right]. \] (7)

This is also called the stretched exponential function.\(^5\) Note that the special case of \( \gamma = 1 \) corresponds to the DRW model, \( \gamma = 0 \) means white noise, and \( \gamma = 2 \) reduces to the Gaussian function.

We further extend our red-noise covariance matrix to account for the QPO phenomenon,

\[ C_{ij}^r = \frac{1}{2} \sigma_i^2 \sigma_j^2 \exp \left[ -\left( \frac{T_{ij}}{\tau_0} \right)^2 \right] \cos \left( \frac{2\pi T_{ij}}{T_q} \right), \] (8)

where \( T_q \) is the period of the QPO. In the case of \( \gamma = 1 \), the power spectral density of the above covariance function becomes

\[ P(f) = \frac{\sigma_i^2 \sigma_j^2}{1 + 4\pi^2 \tau_0^2 (f - f_0)^2} + \frac{\sigma_i^2 \sigma_j^2}{1 + 4\pi^2 \tau_0^2 (f + f_0)^2}. \] (9)

where \( f_0 = 1/T_q \) is the QPO frequency. Note that in the limit of \( T_q \to \infty \), the above equation reduces to Equation (6). In practice, this model is indistinguishable from the DRW model once \( T_q \) is longer than the data span. The noise parameters are \( \theta_n = (\nu, \hat{\sigma}_n, \tau_0, \gamma, T_q) \). We illustrate the red-noise models and discuss their phenomenology in detail in Appendix A.

We use Bayesian model selection to quantify the statistical significance of the presence of periodic signals in quasar light curves. We start with Bayes’ theorem, which states that

\[ P(\theta|d, \mathcal{H}) = \frac{\mathcal{L}(d|\theta, \mathcal{H}) P(\theta|\mathcal{H})}{\mathcal{Z}(d|\mathcal{H})}. \] (10)

Here \( P(\theta|d, \mathcal{H}) \) is the posterior probability distribution function of parameters \( \theta \) given data \( d \) and hypothesis \( \mathcal{H} \); \( \mathcal{L}(d|\theta, \mathcal{H}) \) is the likelihood function given in Equation (1), which describes the probability of observing data given the hypothesis \( \mathcal{H} \) and parameters \( \theta \). Meanwhile, \( P(\theta|\mathcal{H}) \) is the prior distribution of parameters \( \theta \), while \( \mathcal{Z}(d|\mathcal{H}) \) is the Bayesian evidence for hypothesis \( \mathcal{H} \),

\[ \mathcal{Z}(d|\mathcal{H}) = \int d\theta \mathcal{L}(d|\theta, \mathcal{H}) P(\theta|\mathcal{H}). \] (11)

Given the observational data, we wish to compare two hypotheses: \( \mathcal{H}_n \) are the data are consistent with only noise (i.e., \( s = 0 \)) and \( \mathcal{H}_s \) is there is a periodic signal present in the data. The ratio of posterior probability, called the “odds ratio,” between hypotheses \( \mathcal{H}_s \) and \( \mathcal{H}_n \) is

\[ \mathcal{O} = \frac{P(\mathcal{H}_s|d)}{P(\mathcal{H}_n|d)} = \frac{\mathcal{Z}(d|\mathcal{H}_s) P(\mathcal{H}_s)}{\mathcal{Z}(d|\mathcal{H}_n) P(\mathcal{H}_n)}. \] (12)

where \( P(\mathcal{H}_n) \) and \( P(\mathcal{H}_s) \) are the prior probability for hypotheses \( \mathcal{H}_n \) and \( \mathcal{H}_s \), respectively. Assuming equal prior probability for both hypotheses, Bayesian model selection is usually performed by computing the Bayes factor:

\[ B_n^s = \frac{\mathcal{Z}(d|\mathcal{H}_s)}{\mathcal{Z}(d|\mathcal{H}_n)}. \] (13)

In this work, we are concerned with the support for a sinusoidal signal quantified by \( B_n^s \). We also wish to compute the support for the presence of QPO by comparing two models involving Equations (7) and (8) and for the deviation from the DRW model by comparing two models involving Equations (5) and (7). Following Kass & Raftery (1995), the interpretation of Bayes factors is as follows. In a natural logarithmic scale, \( 0 < \ln B < 1 \) indicates that the support for the hypothesis in the numerator is “worth no more than a bare mention,” \( 1 < \ln B < 3 \) implies positive support, \( 3 < \ln B < 5 \) implies strong support, and \( \ln B > 5 \) implies very strong support. These thresholds are somewhat arbitrary; \( \ln B = 8 \) is often adopted as the detection threshold in gravitational-wave astronomy (e.g., Thane & Talbot 2019). Throughout this paper, the log evidence or log Bayes factor refers to the natural logarithm.

3. Application to PG 1302—102

The nearby bright quasar PG 1302—102 has a median V-band magnitude of 15.0 at a redshift of \( z = 0.2784 \). We apply our method to V-band magnitude measurements collected with CRTS (Drake et al. 2009), ASAS-SN (Shappee et al. 2014; Jayasinghe et al. 2019), and LINEAR (Sesar et al. 2011). The CRTS data are publicly available\(^6\) as part of the Catalina Surveys Data Release 2, including 290 photometric measurements taken between 2005 May 6 and 2013 May 30. The ASAS-SN data are downloaded from its photometry database\(^7\) as part of Data Release 9, including 232 measurements acquired between 2013 November 23 and 2018 November 27. The LINEAR data contain 626 measurements made between 2003 January 23 and 2012 June 12. The median measurement uncertainties are 0.06, 0.05, and 0.01 mg for CRTS, ASAS-SN, and LINEAR data, respectively. The span of the combined data set is 15.84 yr. To reduce the computational cost, we average LINEAR data with an interval of 1 day, which reduces the number of data points to 138. The bin size is chosen such that any stochastic time-correlated variations over a timescale greater than 1 day are preserved. The new uncertainty for binned data is computed as the standard deviation of original measurements inside the binning window. The median uncertainty for averaged LINEAR data is 0.015 mag.

Figure 1 shows the data used in our analysis: LINEAR in pink, CRTS in black, and ASAS-SN in blue. The error bars indicate the 1σ measurement uncertainties. We also show the 68% credible region of the sinusoidal signal (gray shaded band) with a period of 5.66±0.10 yr inferred from the data and a red-noise realization (thin light blue line) that contains a QPO with a period of 5.56 yr and might fit the data well.

Before we present the details of our analysis results, we show in Figure 2 the Lomb–Scargle periodogram of PG 1302−102 data (solid blue line), along with the averaged periodogram taken from 10 noise realizations for three models: red noise (dotted red line), red noise plus a QPO (dashed–dotted green line), and red noise plus a sinusoid (dashed cyan line).

---

\(^5\) In the context of describing relaxation in disordered systems, it is called the Kohlrausch–Williams–Watts function.

\(^6\) https://nessi.cacr.caltech.edu/DataRelease/

\(^7\) https://asas-sn.osu.edu/photometry
The vertical black dashed line indicates the data span of 15.84 yr for PG 1302–102. We simulate data using the same sampling in time and error bars as real data to obtain the average periodogram for three models. The following median posterior probability model parameters as inferred from real data are adopted:

1. \( s = -\ln 2.63 \text{ mag yr}^{-2}, t_r = \ln 0.05, \gamma = 0.74 \) for the red-noise hypothesis; 
2. \( s = -\ln 3.47 \text{ mag yr}^{-2}, t_r = \ln 2.94 0, \gamma = 0.52 \) for the red noise + QPO hypothesis; and 
3. \( A = 0.13 \text{ mag}, T_0 = 5.66 \text{ yr}, \phi = 1.64 \text{ rad}, s = -\ln 2.87 \text{ mag yr}^{-2}, t_r = \ln 0.95, \gamma = 0.57 \) for the red noise + sinusoid hypothesis. The purpose of Figure 2 is to provide some insights into how different models might fit the data, instead of a proper accounting of the goodness of these models. The periodogram of real data exhibits a strong peak at 5.6 yr, with secondary peaks at around 1 yr. As one can see, the major peak can be reproduced with both a QPO and a sinusoidal model but not the pure red-noise model. The secondary peaks at around 1 yr are probably artifacts of the irregular sampling of the data, since simulated data that contain no signal at that period also result in apparent peaks there.

### 3.1. Results

To implement the analysis outlined in Section 2 for PG 1302–102 data, we employ the Bilby software package (Ashton et al. 2019), which is a general and versatile Bayesian inference library.
developed primarily for gravitational-wave astronomy. For the stochastic sampling of posteriors and evidence calculation, we use the dynamic nested sampling method developed by Higson et al. (2019), which is available through the Dynasty package (Speagle 2020). The uncertainty of the logarithmic model evidence (\(\ln Z\)) computed by Dynasty is \(O(0.1)\). By performing independent runs of the same sampling process, we verify that the returned values of \(\ln Z\) are consistent within the uncertainty, and that the posterior distributions have converged.

Our priors are specified in Table 1. We consider three cases denoted by prior \(a\), \(b\), and \(c\). The \(a\) prior employs lognormal priors for the DRW model parameters \(\delta^2\) and \(\tau_0\), same as those used in Vaughan et al. (2016), which were based on earlier studies of quasar red noise under the DRW model (Kozlowski et al. 2010; MacLeod et al. 2010; Andrae et al. 2013). The \(b\) prior employs a log-uniform prior for both \(\delta^2\) and \(\tau_0\). Specifically, the prior edges on \(\tau_0\) correspond to 6.7 days and 54.6 yr. For the remaining parameters, we assign uniform priors. There is one mean magnitude \(m\) and one white-noise scaling factor \(\nu\) for each data set (CRTS, ASAS-SN, and LINEAR). Prior \(c\) is the same as prior \(b\), except that we fix these six parameters: \(m\) is fixed at the weighted mean magnitudes 14.98, 15.29, and 15.04 mag for CRTS, ASAS-SN, and LINEAR, respectively, and \(\nu\) is fixed at 0.22, 1.15, and 1.0 for CRTS, ASAS-SN, and LINEAR, respectively, based on analysis of the data under the best-performing hypothesis, as we discuss below. We find that the posteriors on these six parameters are well constrained and do not change with respect to the choice of signal/noise terms included in the hypothesis or the choice of prior for red-noise parameters.

In Table 2, we show the effect of the choice of priors on \(\ln Z\) for the red-noise hypothesis. We find that prior \(b\) results in slightly higher \(\ln Z\) than prior \(a\). We also note that prior \(b\) leads to a posterior distribution of \(\tau_0\) that is in the low-probability tail of the lognormal prior. Therefore, we adopt the less-informative log-uniform prior on red-noise parameters \(\delta^2\) and \(\tau_0\). By fixing \(m\) and \(\nu\), the evidence \(\ln Z\) increases by 14.6 for the full combined data set (comparing prior \(c\) against prior \(b\)). This shows that the inclusion of these parameters is unnecessary, and we only present results based on prior \(c\) in the remaining sections unless otherwise specified.

In comparison to Vaughan et al. (2016), we find that the DRW model is overwhelmingly favored over the sinusoidal model, with \(\ln B = 180 (\log_{10} B = 78)\) using CRTS data alone (under prior \(a\)). This is consistent with the result of \(\log_{10} B > 60\) obtained in Vaughan et al. (2016). However, it may be argued that the choice between pure red noise and pure sinusoid is a false dilemma, as we expect a realistic signal to be characterized by both red noise and a sinusoidal signal. The question investigated here, therefore, is whether or not there is a sinusoidal signal on top of red noise. The corresponding physical picture is that fluctuations in the accretion disk of the primary black hole (or the circumbinary disk) produce red noise and the binary motion results in periodic variations.

In Table 3, we list the log Bayes factor \(\ln B\) for six hypotheses with respect to the red-noise-only hypothesis (Equation (7), denoted as “red”). We examine the presence of a sinusoidal signal (denoted as “sine”) on top of two additional noise scenarios: the DRW model (Equation (5)) and the red noise plus a QPO term (Equation (8), denoted as “red+QPO”). Moreover, \(\ln B^2\) compares the “red+sine” hypothesis to the “DRW+sine” hypothesis and indicates the level of support for deviation from the DRW red noise.

There are a few takeaway messages from Table 3. First, there is very strong evidence against the noise-only hypothesis. Using the full data set (C+L+A), we find strong support for either a sinusoidal signal (\(\ln B = 12.7\)) or a QPO term (\(\ln B = 14.5\)), with the QPO slightly favored with \(\ln B = 1.8\). Assuming an equal prior probability for both signal hypotheses, the QPO hypothesis is favored by an odds ratio of 6.0. While this is inconclusive, we discuss other hints toward the QPO later on. Second, the DRW red-noise model is strongly disfavored, with \(\ln B = 11.4\) (a Bayes factor of \(8.9 \times 10^4\))
using the full data set. Third, there is a moderate reduction in $\ln B$ associated with the inclusion of ASAS-SN data for almost all Bayes factors. For the sinusoidal hypothesis, this is interpreted as evidence against the periodicity by Liu et al. (2018). We look into such a feature in more detail in Section 4. Last, there is no support for the presence of sinusoidal signal on top of the red+QPO scenario.

The full posterior distribution of model parameters for the best-performing hypothesis (red+QPO, using prior $b$) is presented in Appendix B. We find that the CRTS measurement uncertainties are significantly overestimated by a factor of 5 ($\nu = 0.20^{+0.01}_{-0.01}$), whereas those of ASAS-SN are slightly underestimated ($\nu = 1.16^{+0.07}_{-0.06}$). The unusual value of $\nu$ for CRTS data is consistent with that found in Vaughan et al. (2016) and can be verified by binning the original measurements inside a 1 day window and computing the dispersion. For LINEAR data, we find $\nu = 0.91^{+0.12}_{-0.11}$, which is consistent with 1.

Figure 3 shows the posterior distributions for the red+QPO (left panel) and red+sine (right panel) hypotheses for three combinations of data sets. It can be seen that the use of the full data set produces long-term correlations in the data, meaning the noise component in new data is not independent from old data. To demonstrate this effect, we inject a periodic signal into 10 random realizations of quasar red noise. We choose exactly the same sampling and error bars as shown in Figure 1. Each of the 10 data sets contains three subsets, from LINEAR, CRTS, and ASAS-SN. We compute the Bayes factor, comparing the red+sine hypothesis against the red-noise-only hypothesis, for different combinations of subsets.

Comparing the two panels in Figure 3, we find the following. (1) The posteriors for the red+QPO hypothesis are better constrained. (2) There appear to be two posterior modes for the red+sine hypothesis for the red-noise amplitude $\hat{\sigma}^2$ and timescale $\tau_0$. These two parameters are correlated, and one of the two modes—large $\tau_0$ and small $\hat{\sigma}^2$—is consistent with the posterior distribution found under the red+QPO hypothesis. (3) The posteriors for three combinations of data sets are consistent for the red+QPO hypothesis, whereas those for the red+sine hypothesis are unstable. Specifically, the credible region for the sinusoidal period $T_0$ is inconsistent between CRTS (4.54$^{+0.16}_{-0.13}$ yr) and the full data set (5.66$^{+0.17}_{-0.10}$ yr). This comparison is an example of what is known as a posterior predictive check, in which one performs sanity tests to make sure the models are suitable for Bayesian inference. The instability of posterior distributions with the inclusion of more data may suggest that the red+sine hypothesis does not fully describe the data.

4. Discussion

4.1. What Does It Mean that the Significance of the Periodicity in PG 1302−102 Goes Down When We Add More Data?

The growth in detection significance of a sinusoidal signal is not guaranteed in the presence of red noise. A red-noise process produces long-term correlations in the data, meaning the noise component in new data is not independent from old data. To demonstrate this effect, we inject a periodic signal into 10 random realizations of quasar red noise. We choose exactly the same sampling and error bars as shown in Figure 1. Each of the 10 data sets contains three subsets, from LINEAR, CRTS, and ASAS-SN. We compute the Bayes factor, comparing the red+sine hypothesis against the red-noise-only hypothesis, for different combinations of subsets.

We choose the following parameters, which are consistent with posterior estimates of PG 1302−102: $A = 0.12$ mag, $T_0 = 5.53$ yr, $\phi = 5.0$ rad, $\ln \hat{\sigma}^2 = -2.56$ mag$^2$ yr$^{-1}$, $\ln(\tau_0/$yr$) = 0.15$, $\gamma = 0.62$, $\nu_{\text{CRTS}} = 0.21$, $\nu_{\text{LINEAR}} = 0.96$, and $\nu_{\text{ASAS-SN}} = 1.15$. The red-noise realization is generated using $n = L r$. Here $L$ is a lower triangular matrix obtained from the Cholesky decomposition of the noise covariance matrix $C = L L^T$, where $L^T$ is the conjugate transpose of $L$. The $N \times 1$ vector of $r$ contains $N$ independent random numbers that follow the standard normal distribution.
Figure 4. Log Bayes factor (ln $B$) that compares the red noise plus a sinusoidal signal hypothesis against the noise-only hypothesis for different combinations of data sets: C—CRTS, A—ASAS-SN, L—LINEAR. Blue diamonds are for real data, whereas light blue (gray) dots are for simulated data that contain a sinusoidal signal (pure red noise). The red (black) horizontal line marks $\ln B = 8(0)$.

Figure 4 shows the Bayes factors from 10 simulated light curves of PG 1302–102 as light blue dots with green lines. Several features are noteworthy. First, the general trend is that Bayes factors grow when we add LINEAR and ASAS-SN data to CRTS data. Second, there are two noise realizations where the addition of LINEAR data onto CRTS data results in a reduced detection significance of periodicity. We note that these two realizations have relatively low initial Bayes factors ($\ln B \lesssim 6$). Therefore, we conclude that (1) whether or not the periodicity significance grows with the inclusion of a certain set of additional data cannot be used as a simple criterion for true periodicity, and (2) the Bayes factor is more likely to grow with time once the initial Bayes factor is high (e.g., $\ln B \gtrsim 8$). Third, the addition of ASAS-SN data generally leads to a greater improvement in detection significance than LINEAR. As a sanity check, we also show the log Bayes factors for 10 noise-only data sets as gray dots in Figure 4, using the same noise parameters mentioned earlier. In this case, the log Bayes factors are around $-2$, correctly disfavoring the presence of sinusoidal signals. The addition of new data does not change the Bayes factor significantly.

In Figure 4, we also show results from the analysis of real data as blue diamonds. The evolution of $\ln B$, particularly the reduction in $\ln B$ associated with ASAS-SN data, is in drastic contrast with our simulations. In Figure 5, we plot the posterior distributions for one of the sinusoidal injections for different combinations of subdata sets. In comparison to the right panel of Figure 3, which is the counterpart plot for real data, we find that the posteriors are more stable. Taken together, our simulations hint toward a problem with the sinusoidal signal hypothesis for PG 1302–102.

Lastly, as a side note, the simulation study presented here highlights the challenge in the periodicity search in quasar light curves; namely, there is only one realization of the red noise.

This situation is in contrast with the search for continuous gravitational waves (also essentially sinusoidal signals) using pulsar timing arrays. Many millisecond pulsars in the timing array have been found to exhibit red noise for which the origin is largely unknown. However, these red-noise processes, if intrinsic to pulsars themselves, are not expected to correlate among different pulsars. A false detection caused by red noise in one particular pulsar can be ruled out by cross-checking other pulsar data. This sort of cross-check is not always possible when searching for binary black holes in quasar light curves.

4.2. Can We Distinguish a Sinusoid from a QPO?

Our analysis of PG 1302–102 data reveals modest evidence for a QPO over a sinusoidal signal. It is natural to ask a question: under what circumstances can we distinguish a true sinusoid from a QPO? To find out, we compute the Bayes factor between the sinusoidal and QPO hypotheses for simulated data that contain a sinusoidal signal and some red noise. We take the sampling and error bars of PG 1302–102 data and vary the signal period so that the data (spanning 15.84 yr) include three, six, and nine wave cycles. For injection $a$, we choose the following parameters: $A = 0.13$ mag, $\phi = 5.0$ rad, $\ln \sigma^2 = -2.56$ mag$^2$ yr$^{-2}$, $\ln(\sigma_0/\gamma) = 0.15$, and $\gamma = 0.62$. Injection $b$ is the same except that we increase the signal amplitude ($A = 0.2$ mag) and reduce the red-noise amplitude ($\ln \sigma^2 = -4.56$ mag$^2$ yr$^{-1}$).

Table 4 lists the log Bayes factors for both injections. One can see that the Bayes factor that supports the presence of a sinusoidal signal increases with the number of signal cycles, as expected. However, for injection $a$, the sinusoid is indistinguishable from a QPO, even after nine wave cycles. In fact, the Bayes factor slightly favors the QPO hypothesis for the six- and nine-cycle cases. For the stronger injection $b$, it becomes possible to tell that the signal is a sinusoid instead of a QPO,
Figure 6. Posterior distributions of the red+QPO and red+sine hypotheses for a simulated data set that contains a sinusoidal signal on top of red noise (injection a, the three-cycle case in Table 4). Black dashed lines mark the true values of the parameters.

and the corresponding Bayes factor grows with the number of wave cycles.

Figure 6 shows the posterior distributions for the three-cycle case of injection a. Whereas the injection parameters are correctly recovered for the true red+sine hypothesis, red-noise parameters are mistaken by the (incorrect) red+QPO hypothesis. Noting that the quality factor of the QPO is given by $\pi \tau_0 / T_p$, it is unsurprising that the QPO hypothesis results in posteriors that peak at large $\tau_0$ and small red-noise amplitude $\tilde{\sigma}$.

4.3. How Does Binning Affect the Detection Significance and Parameter Estimation?

It is common that some sort of averaging or binning is applied when analyzing time-series data. Information is inevitably lost in this process. Here we demonstrate the potentially adverse effect of binning on the periodicity detection significance and parameter estimation precision with an example. We choose the simulated data set that gives rise to the highest log Bayes factor in Figure 4. Because the LINEAR data used in this work have already gone through an averaging process with an interval of 1 day, we only use simulated data for CRTS and ASAS-SN. Figure 7 shows the simulated data, their binned versions with an interval of 100 days, and the injected sinusoidal signal.

We compute two Bayes factors: (a) the red+sine hypothesis against the red-noise-only hypothesis and (b) the red+sine hypothesis against the DRW+sine hypothesis. Cases (a) and (b) indicate the level of support for periodicity and deviation from the DRW model, respectively. The log Bayes factors are 13.0 and 6.3 for cases (a) and (b), respectively, using the original data. These Bayes factors become 7.7 and $-0.1$ for cases (a) and (b), respectively, using the binned data. Therefore, the binning process reduces our ability to not only detect a periodicity but also identify deviation from the DRW red-noise model (noting the high-frequency red-noise fluctuations illustrated in Figure 1).

Table 4

| Number of Signal Cycles | Three | Six  | Nine |
|-------------------------|-------|------|------|
| Injection a             | $\ln B_{\text{red}}^{\text{true}}$ | 15.0 | 25.5 | 36.4 |
|                         | $\ln B_{\text{red+QPO}}$         | 0.1  | $-2.4$ | $-2.1$ |
| Injection b             | $\ln B_{\text{red}}^{\text{true}}$ | 40.6 | 68.9 | 101.3 |
|                         | $\ln B_{\text{red+QPO}}$         | 4.6  | 7.5  | 13.3 |

Note. Results are listed for three different numbers of signal cycles. Injection b contains a higher-amplitude signal than Injection a.

4.4. On the Implementation of Our Method to a Large Sample of Quasar Light Curves

The computational cost in our analysis is dominated by the computation of the inverse and determinant of the $N \times N$ noise covariance matrix associated with the likelihood evaluation, which scales as $N^3$. For the entire data set of PG 1302–102 analyzed in this work, $N = 660$. A single likelihood evaluation takes on the order of 0.1 s, and the full parameter estimation and evidence calculation for such a data set take about 24 hr on a single modern CPU core. The computational cost is the reason we choose to average LINEAR data in an interval of 1 day, which reduces the total number of data points from 1148 to 660. Furthermore, there are a large number of quasars for which long-term photometric measurements are available, for example, $10^7$ quasars from the CRTS, as processed in Graham et al. (2015a). Therefore, further speed-up of our method is highly desirable.

This is a well-known problem in the analysis of astronomical time series. In pulsar timing arrays, there are normally hundreds to thousands of times of arrival measurements for each pulsar, and the problem involves correlation analysis with data from dozens of pulsars. Various acceleration/approximation techniques have been proposed to enable full Bayesian analysis of pulsar timing array data. One popular method is to approximate the red noise as the sum of $k$ Fourier components; thus, its covariance matrix is given by $C = F \Phi F^T$, where $F$ is $N \times k$ (with $k \ll N$) and $\Phi$ is a $k \times k$ diagonal matrix. This turns the computationally heavy inversion of $C$ into the lower-rank diagonal matrix inversion $\Phi^{-1}$ through the Woodbury matrix lemma (e.g., Lentati et al. 2013; van Haasteren & Vallisneri 2015). Elsewhere, Foreman-Mackey et al. (2017) proposed a fast Gaussian process model that makes use of the feature where the covariance function can be expressed as a...
mixture of complex exponentials. Implemented as the Cerite package, its computational cost scales linearly with the size of the data set; the same scaling for computational cost is shared by the CARMA method (Kelly et al. 2014). All of these methods may prove useful for the analysis of light curves for a large number of quasars within the Bayesian framework developed here.

5. Conclusions

Subparsec supermassive binary black holes are crucial in our understanding of galaxy evolution. They are also the primary sources of the nanohertz gravitational waves highly anticipated by international pulsar timing array campaigns. While nearly impossible to resolve through direct imaging, such close binaries are expected to produce periodic variations in the light curves of active galactic nuclei. New candidates of periodicity are reported on a nearly monthly basis.

In this work, we propose a fully Bayesian method for the identification of periodicity in astronomical time series that exhibit red noise. We apply this method to one of the most promising periodicity candidates, PG 1302−102, using data from CRTS, ASAS-SN, and LINEAR. Our main findings are as follows.

1. There is very strong support for the presence of either a sinusoidal signal (\(\ln B = 12.7\)) or a QPO (\(\ln B = 14.5\)) at a period of 5.6 yr. The data slightly favor the QPO hypothesis with an odds ratio of 6.
2. The inclusion of ASAS-SN data reduces the log Bayes factor that supports the presence of a sinusoidal signal by 1.6. This, combined with the fact that the posterior distribution of the sinusoidal period is unstable when we combine new data with CRTS, provides further evidence against the sinusoidal hypothesis.
3. There is also conclusive evidence for deviation from the DRW red-noise model, with \(\ln B = 11.4\). The noise power spectral density is shallower than a power law with an index of 2, and the red-noise damping timescale is long, \(\gtrsim 10\) yr.

We perform simulations of sinusoidal signals and red noise and demonstrate the following.

1. The growth of the periodicity significance with additional data is not guaranteed because of the stochastic fluctuations of red noise. For data sets like LINEAR, CRTS, and ASAS-SN, the log Bayes factor is expected to grow once \(\ln B \gtrsim 8\) is achieved with initial data.
2. Our method is capable of distinguishing a sinusoidal signal from a QPO. However, this is only possible for strong signals or a large number of wave cycles. A strong prior constraint on the red noise would also help.
3. The use of data binning can reduce our ability to detect periodicity or deviation from the DRW model because the binning throws away high-frequency information that helps estimate the spectral slope of the red noise (see Figure A1).

We show that the data of PG 1302−102 favor the presence of a quasiperiodicity at around 5.6 yr against the noise hypothesis with a Bayes factor of \(2 \times 10^6\). However, given that PG 1302−102 was selected as the most periodic candidate out of \(2 \times 10^5\) quasars, it cannot be ruled out that such a quasiperiodicity is caused by pure red noise. Assuming that it is indeed a QPO and the central black hole is \(10^{8.5} M_\odot\) (Graham et al. 2015b), we infer a period corresponding to a Keplerian orbit at 344 Schwarzschild radii (\(\approx 0.01\) pc). We note that the QPO found in PG 1302−102 roughly corresponds to the low-frequency QPOs seen in stellar-mass black hole X-ray binaries, although the QPO frequency of PG 1302−102 lies 1 order of magnitude below the linear scaling relation between black hole mass and QPO frequency fitted to several candidates in Smith et al. (2018). The physical origins of QPOs are poorly understood; therefore, such a scaling relation might not exist if there are different mechanisms driving QPOs in small and

![Figure 7. Simulated light curve of PG 1302−102 that includes an injected sinusoidal signal. Mean magnitudes have been subtracted. Filled circles are original data, whereas open circles are binned data with an interval of 100 days.](image)
big black holes. Nevertheless, the QPO hypothesis of PG 1302 −102 can be tested with continued observations, and if the quasiperiodicity signature is confirmed, it may have significant implications for the accretion physics of supermassive black holes.

Finally, our Bayesian framework can be adopted to establish unambiguous binary black hole detections with the following extensions.

1. Apply this method to various physical models for the periodicity, such as the periodic mass accretion rate of the binary (e.g., Farris et al. 2014) or jet precession of a single or binary supermassive black hole (e.g., Abraham & Carrara 1998; Kudryavtseva et al. 2011; Britzen et al. 2018), in addition to relativistic Doppler boosting.

2. Use more sophisticated signal models that account for a cold spot–induced perturbation in the accretion disk for PG 1302−102 (Kovačević et al. 2019) or post-Newtonian orbital evolution for OJ 287 (Dey et al. 2018).

3. Combine multiwavelength observations (e.g., Xin et al. 2020) and other signatures associated with a binary black hole, for example, Doppler velocity offsets in broad emission line profiles (Bon et al. 2012; Eracleous et al. 2012; Li et al. 2016) and flux deficits in the spectral energy distribution (Gültekin & Miller 2012; Guo et al. 2020). See Zheng et al. (2016) for a binary black hole candidate for which different types of signatures are analyzed.

4. Use astrophysically motivated population priors. For example, the period distribution of the binary black hole population is expected to be dominated by long periods, and the distribution of red-noise parameters can be obtained by applying our method to a large number of active galactic nuclei using hierarchical inference.

We thank the anonymous referee for helpful comments that encouraged us to do a more thorough study of PG 1302−102. We thank Tingting Liu for sharing the PG 1302 data used in Liu et al. (2018) and Boris Goncharov and Yuri Levin for useful discussions. This work is supported by ARC CE170100004 and ARC FT150100281. The Catalina Sky Survey is funded by NASA under grant No. NNG05GF22G issued through the Science Mission Directorate Near-Earth Objects Observations Program. The CRTS survey is supported by the U.S. National Science Foundation under grants AST-0909182 and AST-1313422. We use the ChainConsumer package (Hinton 2016) for some of the plots presented here. Codes and data to reproduce the major results of this work are publicly available at https://github.com/ZhuXJ1/SuperBayes.

Appendix A
Red-noise Models

In Figure A1, we show the covariance function (left panel) and power spectral density (right panel) for the red-noise models considered in this work. Black dashed lines are for the DRW model, i.e., $\gamma = 1$, and colored lines are for different values of $\gamma$. Additionally, we show the case of a QPO on top of the DRW red noise with a green line. The gray vertical line in the right panel indicates $1/T_{\text{obs}}$, where $T_{\text{obs}} = 15.84$ yr is the data span of PG 1302−102. The timescale $T_0$ and the QPO period $T_Q$ are both set to be 1 yr in these examples.

The prevalent DRW red-noise model features a spectral turnover at a frequency that corresponds to the damping timescale $T_0$ and a power-law spectral density with a power-law index of 2 at high frequency ($f \gg 1/T_0$). As can be seen in the right panel of Figure A1, our general red-noise model covers a broad range of power-law spectral shapes, from a power-law index of zero ($\gamma = 0$) to an extremely steep spectrum ($\gamma = 2$). In order to measure the spectral shapes of the quasar red noise, i.e., the $\gamma$ parameter, the high-frequency part is critical, since that is where $\gamma$ has the largest influence. In the case that $T_0 \ll T_{\text{obs}}$, various models reduce to the power-law model. The QPO model, on the other hand, is drastically different, as it features a peak in the power spectral density. The quality factor of the QPO, defined as the ratio of the peak frequency of the QPO to its width, is determined by $\pi T_0/T_Q$.

We note that the DRW model and the QPO prescription adopted here are special cases of the flexible continuous-time autoregressive moving average (CARMA) models. Kelly et al. (2014) demonstrated that CARMA models are useful in identifying new features in the power spectral density of astronomical

---

**Figure A1.** Covariance function $C(\Delta \tau)$ (left) and power spectral density $P(f)$ (right), both in arbitrary units, for different values of $\gamma$ (see Equation (7)) and the presence of a QPO on top of the DRW red noise ($\gamma = 1$; see Equation (8)). The gray vertical line in the right panel corresponds to the data span of 15.84 yr for PG 1302−102.
time-series data. Since the power spectral density of CARMA models can be expressed as a sum of Lorentzian functions, in a similar fashion to the Celerite model described in Foreman-Mackey et al. (2017), it allows fast evaluation of the likelihood function, which scales linearly with the number of data points. This makes it particularly powerful for analyzing massive time-domain data from a large number of objects. We leave the exploration of the CARMA and Celerite models in our Bayesian inference framework for future work.

Appendix B

Posterior Distributions from Analysis of Real Data

Here we present full posterior distributions from the analysis of the combined data set (Figure 1) from CRTS, LINEAR, and ASAS-SN for PG 1302–102. Figure B1 shows the distributions for the red noise plus QPO hypothesis, which is the favored hypothesis given the data. Note that the mean offset parameters $m$ for LINEAR and CRTS data are highly covariant because there is a significant overlap in time for the two.

**Figure B1.** Posterior distributions of parameters for the red noise + QPO hypothesis for the light curve of PG 1302–102.
Appendix C
Posterior Distributions from Analysis of Simulated Data

Here we present full posterior distributions from the analysis of a simulated data set (Figure 7) that includes an injected sinusoidal signal. Figure C1 shows the distributions for the original data, whereas Figure C2 shows the results for the binned data with an interval of 100 days.

Figure C1. Posterior distributions of parameters for the sinusoidal signal plus red noise hypothesis for a simulated data set of PG 1302−102 (shown in Figure 7), which included an injected signal. Orange lines mark the injection values.
**References**

Abraham, Z., & Carrara, E. A. 1998, *ApJ*, 496, 172
Aggarwal, K., Arzoumanian, Z., Baker, P., et al. 2019, *ApJ*, 880, 116
Andrae, R., Kim, D. W., & Bailer-Jones, C. A. L. 2013, *A&A*, 554, A137
Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, *ApJS*, 241, 27
Babak, S., Petitjean, A., Sesana, A., et al. 2016, *MNRAS*, 455, 1665
Bansal, K., Taylor, G. B., Peck, A. B., Zavala, R. T., & Romani, R. W. 2017, *ApJ*, 843, 14
Begelman, M. C., Blandford, R. D., & Rees, M. J. 1980, *Natur*, 287, 307
Bon, E., Jovanović, P., Marziani, P., et al. 2012, *ApJ*, 759, 118
Britzen, S., Fendt, C., Witzel, G., et al. 2018, *MNRAS*, 478, 3199
Charisi, M., Bartos, I., Haiman, Z., et al. 2016, *MNRAS*, 463, 2145
Chen, S., Sesana, A., & Conselice, C. J. 2019, *MNRAS*, 488, 401
Colpi, M. 2014, *SSRv*, 183, 189
Cordes, J., & McLaughlin, M. A. 2019, *BAAS*, 51, 447
Dey, L., Gopakumar, A., Valtonen, M., et al. 2019, *Univ*, 5, 108
Dey, L., Valtonen, M. J., Gopakumar, A., et al. 2018, *ApJ*, 866, 11
D’Orazio, D. J., Haiman, Z., & Schiminovich, D. 2015, *Natur*, 525, 351
Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, *ApJ*, 696, 870
Eracleous, M., Boroson, T. A., Halpern, J. P., & Liu, J. 2012, *ApJS*, 201, 23

**Figure C2.** Same as Figure C1 but for the binned data with an interval of 100 days.
