Tools for period searching in AGN in the era of “Big Data”

Saikruba Krishnan[1], Alex Markowitz[1,2] and Aleksander Schwarzenberg-Czerny[1]

1. Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, Bartycka 18, 00–716 Warsaw, Poland
2. Center for Astrophysics and Space Sciences, Univ. of California, San Diego, MC 0424, La Jolla, CA 92093-0424, USA

Active Galactic Nuclei (AGN) persistently emit across the electromagnetic spectrum and are dominated by stochastic, aperiodic emission that is variable on timescales from hours to decades. Because pure stochastic red noise signals spuriously mimic few-cycle periods, the presence of the stochastic variability impedes our ability to robustly confirm if any periodic signals are present in AGN. In addition, we have already entered the era of “Big Data,” with current and near-future large-area monitoring programmes such as LSST facilitating data trawls for periodicities; developing the proper know-how for period searching is thus essential. Hence in our project we try to account for the presence of red noise using various methods — Auto-correlation Function (ACF), Phase dispersion minimization (PDM), wavelet analysis and Bayesian analysis — and test if each method can robustly distinguish between pure red-noise processes and mixtures of a strictly- or quasi-periodic signal (QPO) with red noise, while pursuing the following questions: When the variability process is pure red noise (no QPO present), what is the false-alarm probability and how does it depend on broadband continuum power spectral density (PSD) shape? When there is intrinsically a mixture of red noise and a QPO, is there a range in detection sensitivity between the various methods as a function of broadband spectral slopes and different strengths of QPO against the red noise? Here we present some preliminary results from analysis in progress using the ACF and the PDM. We compare our results to models of binary supermassive black hole (SMBH) systems to constrain the regions of parameter space where detection of periodicities due to binary orbital motion against red noise from the accretion disk is feasible with typical ground-based monitoring programs.

1 Introduction

Active Galactic Nuclei (AGNs) are supermassive black holes (SMBHs) actively accreting gas and residing at the centers of large galaxies. They persistently release large amounts of light across the electromagnetic spectrum as well as collimated jets of outflowing gas. AGN can thus provide tests of General Relativity via probes of the dynamical accretion flows in the regime of extreme gravity and in cases where two SMBHs are part of a gravitationally-bound binary system. Variability analysis of AGNs’ continuum flux enables us to search for characteristic time scales which in turn provide insight into how accretion processes and variability mechanisms depend on system parameters such as the black hole mass $M_{\rm BH}$, luminosity, accretion rate relative to Eddington, etc.

In many stellar-mass black hole X-ray binaries (BHXB), power spectral density (PSD) functions reveal quasi-periodic oscillations (QPOs; e.g., Wijnands et al. 1999;

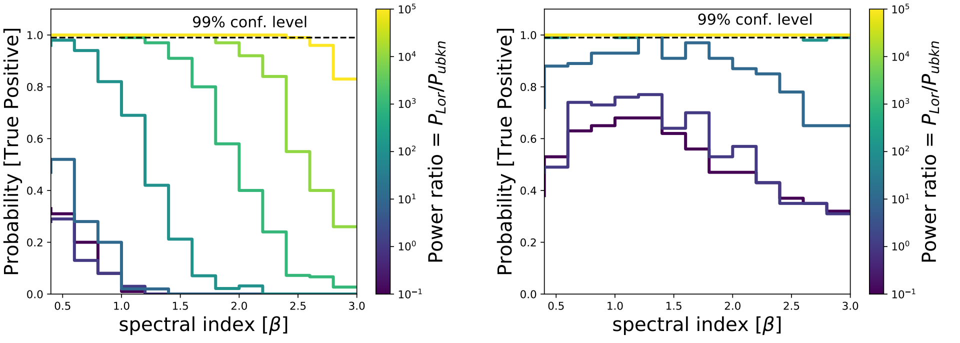
Casella et al. 2004; Motta et al. 2015). They may originate via Lense-Thirring precession in the inner disk (e.g., Ingram & Done 2012), and/or various oscillation modes and resonances (e.g., Perez et al. 1997; Abramowicz et al. 2003). We may expect similar processes to occur in the disks of Seyfert AGN, given the evidence for identical accretion processes in both classes from radio/X-ray luminosity relations (Merloni et al. 2003) and X-ray PSDs (Edelson & Nandra 1999; Uttley et al. 2002; Markowitz et al. 2003). In addition, claimed periodicities in blazars can be interpreted as due to precessing jets (e.g., Villata & Raiteri 1999; Li et al. 2009; Sandrinelli et al. 2016). Recently, claimed periods have been ascribed to gravitationally-bound SMBH binaries, wherein streams of circumbinary gas get periodically modulated (Lehto & Valtonen 1996, Valtonen et al. 2006, Graham et al. 2015, Liu et al. 2015, Charisi et al. 2016). In AGN, data quality (e.g., frequency resolution) is usually insufficient to confirm QPOs using the PSD (Vaughan & Uttley, 2005); there have been only $\sim$ a couple robust claims so far (Gierliński et al. 2008; Alston et al. 2019). QPO detection in AGN usually thus relies on alternate methods such as the ACF, PDM, or sinusoidal fitting. However, pure stochastic red noise processes can spuriously mimic few-cycle sinusoid-like periods (Kozłowski et al. 2010; Vaughan et al. 2016). Many claims of AGN periodicities in the literature are few-cycle, and/or the authors did not account for the red noise background, making it likely that in those cases, the claimed periodicity does not exist, and is merely an artefact of red noise.

The situation is becoming more urgent because we have already entered the era of "Big Data" given the fact that ground-based observing programmes such as PanSTARRS, the Palomar Transient Factory, and LOw Frequency ARray (LOFAR) and future programmes like LSST, ZTF and SKA now monitor or will monitor large fractions of the sky. The resulting light curve databases allow or will allow period searches over 10^3 to 10^6 AGN simultaneously.

The goal of our project is to formulate guidelines on the proper use of several commonly-used statistical methods such as the auto-correlation function (ACF), phase dispersion minimization (PDM), wavelets, and CARMA for searching for deterministic signals in red noise-dominated AGN light curves and to test their efficacy in robustly distinguishing between a pure stochastic red noise process (no QPOs) and a mixture of red noise and a QPO signal. In these proceedings we share some of the preliminary results on using the ACF and PDM.

2 Methodology

We use Monte Carlo simulations to empirically test these methods starting with a range of input model PSD shapes. We assume a simple unbroken power-law model for the broadband red-noise, and add QPOs described in frequency space by a Lorentzian profile whose width is quantified by the quality factor Q. We choose Lorentzians with a relatively high quality factor, $Q = 30$. For initial tests, we simulate "ideal sampling": evenly-sampled light curves 250 days long having one point per day, produced using the Timmer & Konig (1995) algorithm, and similar to ground-based monitoring program with a sun gap. We do not include any Poisson noise in these initial tests. The analyses are done for three representative test Lorentzian frequencies: a low-frequency QPO ($\sim$ 4 cycles) of 2.0×10^{-7} Hz, a mid-frequency QPO ($\sim$ 17 cycles) of 8.0×10^{-7} Hz, and a high-frequency QPO ($\sim$ 97 cycles) of 32.0×10^{-7} Hz, and we determine the detection thresholds for QPOs



(a) without filtering (b) After filtering

Fig. 1: Results for the ACF: We plot here the probability of detecting the first harmonic of the periodic signal in the expected frequency range, as a function of power ratio $P_{\rm rat}$ and red noise PSD power-law slope β. We plot here results for the mid-frequency QPO, for (a) unfiltered and (b) filtered light curves.

when mixed with broadband red noise and also at different ranges of power ratio ($P_{\rm rat}$), defined as the ratio of QPO peak power to the power of the red noise power law at that frequency.

2.1 Test for QPOs mixed with red noise using the ACF

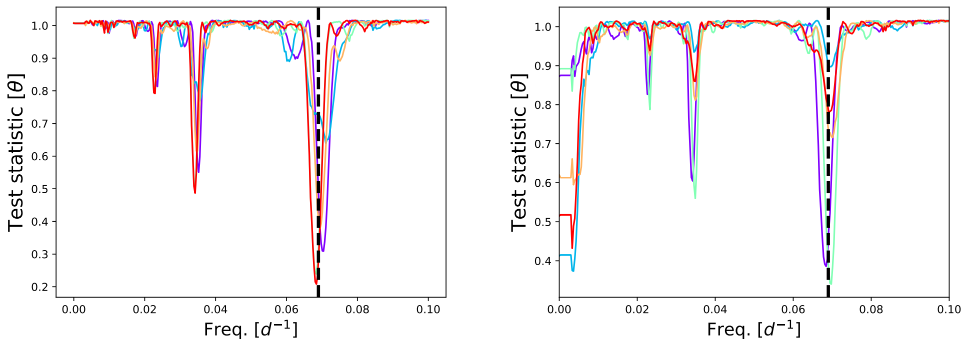
In the "ideal" case of a QPO with no red noise present, the ACF is cosine-like, with the time delay of the first major peak (the first harmonic, here called "T4") occurring within e.g., ±20% of the frequency centroid (99% confidence) for the mid-frequency case. We perform the Monte Carlo simulations across a range of red noise power-law slopes $\beta = 0.4$–3.0, and power ratios $P_{\rm rat}$ ranging from 10^{-1} to 10^5, and we register a true positive detection only when T4 falls in the expected time range.

The detection rate decreases as the power ratio decreases and as red noise PSD slope β increases: there is in fact no significant detection when $P_{\rm rat}$ is $\lesssim 10^2$ and $\beta \gtrsim 0.6$ (Fig. 1). In the case of the mid-frequency QPO, we conclude that significant detection of the first and second harmonics with 99% reliability occurs only for $P_{\rm rat} \gtrsim 10^3, 10^4$, and 10^5 for $\beta \lesssim 1.2, 1.8$, and 2.6 respectively.

2.2 Test for QPOs mixed with red noise using PDM

The Phase dispersion minimization (PDM) method can be used for searching for non-sinusoidal pulsations (Stellingwerf, 1978). The test statistic θ parameter of the PDM, which follows a beta distribution (Schwarzenberg-Czerny, 1997), gives the measure of fit quality between the observations & the model at each test frequency, with a deep minimum indicating the frequency corresponding to a periodic/QPO signal.

We again first examined the ideal case of a pure Lorentzian with $Q = 30$ and with no red noise for each representative test frequency, to delineate the region of frequency–θ space where the "ideal" signal occurs (Fig. 2a). We then simulated light curves as covering the same ranges of red noise power-law slope and values of $P_{\rm rat}$ as for the ACF. Again, the true positive detection rate decreases as the $P_{\rm rat}$ decreases and/or as red noise slope increases. The presence of red noise spuriously causes low values of the minimum value of the test statistic, $\theta_{\rm min}$, to occur, preferentially towards the lowest few frequency bins, and for a significant fraction of realizations; one such case is illustrated in Fig. 2b. Overall, we find that significant detections of the QPO occur only for $P_{\rm rat} \sim 10^5$ when β is < 2. As a general warning, the red noise (especially with relatively steeper slopes, $\beta \gtrsim 2.0$) spuriously causes significant fraction of the test statistic minima of PDM to occur towards the few lowest frequency bins for $P_{\rm rat} \lesssim 10^4$.



(a) (b)

Fig. 2: The distribution of test statistic value θ of PDM against frequency (a) for a pure Lorentzian, here, a mid-frequency QPO signal with $Q = 30$; (b) for a mid-frequency QPO signal mixed with red noise with spectral slope of $\beta = 2.6$, and power ratio $P_{\rm rat} = 10^4$.

3 Reducing the effect of red noise processes

We are in the process of exploring whether simple light curve filtering techniques are effective in removing the effects of red noise while preserving putative periodic signals. We bin the light curve using a moving average, which is then subtracted. Given that users will not know in advance the period of the putative periodic signal (if it even exists), we recommend that users simply try a range of bin sizes. Applying the ACF to filtered light curves results in the case of the mid-frequency QPO, we find that periods are significantly detected with $\geq$99% reliability across all PSD red noise slopes for $\log(P_{\rm rat}) \gtrsim 2$ — two orders of magnitude lower compared to not filtering. In the case of the PDM, we find significant detections for the mid-frequency for $\log(P_{\rm rat}) \gtrsim 3$ for all slopes. There are thus strong indications that filtering leads to significant improvement in sensitivity in $P_{\rm rat}$. A warning, however, is that if the periodic signal strongly dominates over the red noise; that component's variability would be removed by this technique. We are in the process of defining the $P_{\rm rat}$ regime where this technique is optimal.

4 Application to SMBHB systems & realistic monitoring programs

In SMBHB systems, the binaries' orbital motion periodically modulates accretion streams infalling from the inner edge of the circumbinary gas disk; we can test the feasibility of detecting this periodic variability against the variability from disk accretion by constructing model power spectra. The PSD components are broadband red noise from variability in the mini-disks (see e.g., Markowitz et al. 2003; Arévalo & Uttley 2006; McHardy et al. 2006) and QPOs following hydrodynamical simulations (Farris et al. 2014), and we consider ranges of total masses, mass ratios, binary separations. Tentative results considering the temporal frequencies probed by typical multi-year ground-based monitoring programs indicate that detection is not feasible for most SMBHB systems; detection is likely only for systems with total masses $\gtrsim 10^{7.5}\,\mathrm{M}_{\odot}$ and separation distances $\lesssim 10^{2}$ gravitational radii. For all other cases, the periodic signal is overwhelmed by disk variability at that frequency and/or falls outside the range of temporal frequencies probed.

5 Conclusions

- The significance of detection of the QPO signal when it is mixed with red noise depends on both the strength of the QPO signal against the unbroken power law PSD slope and also the steepness of the spectral slope. In order to ensure a 99% or higher reliability for detection, one typically requires a high power ratio — in our mid-frequency case, power ratios were required to be $\sim 10^{5}$, 10^{4}, and 10^{3} for red noise power-law slopes of ~ 2.6, 1.8, and 1.2, respectively.

- Consequently, if one claims presence of a quasi- or strictly-periodic signal using the ACF or PDM, then they are inherently making a claim of about the intrinsic RMS strength of the period relative to that of the red noise.

- A filtering technique consisting of subtracting a smoothed component of the light curve improves the probability of detection of the period: typically, we find that with filtering, one can reliably detect QPOs with power ratios $\sim$2 orders of magnitude lower compared to not filtering across all tested values of PSD slope β, for the mid-frequency QPO case.

- We are in the process of testing sampling patterns corresponding to actual monitoring programs (e.g, LSST, OVRO) which may feature annual Sun gaps and/or uneven sampling. Results from these tests will be discussed in more detail in our planned paper.

- Tentatively, typical monitoring programs covering timescales of days to several years will likely be able to detect periodic signals from SMBH binary systems likely only if the binaries' separation is $\lesssim 10^{2}$ gravitational radii and their total mass is $\gtrsim 10^{7.5}$ $\mathrm{M}_{\odot}$.

- Work is in progress on test the weighted wavelet-z transform (WWZ) method for separating red noise and QPOs: The wavelet's ability to deconvolve signals into both temporal frequency and time enhances the number of independent trials and increases the "look elsewhere" effect; the presence of red noise, given its ability to mimic few-cycle sinusoids, can thus create many false-positive power peaks. Very tentatively, we find that one must consider regions of power lying completely above threshold factors of 20–30 above the local average power (depending on local average power spectral slope) to filter out potential spurious signals.

- Work on using CARMA, the Bayesian time-domain fitting method, is in progress.

Acknowledgements. S.K., A.M., and A.S.-C. all acknowledge support from NCN grant 2016/23/B/ST9/03123.

References

Abramowicz, M. A., Bulik, T., Bursa, M., Kluźniak, W., *A&A* **404**, L21 (2003)

Alston, W. N., et al., *MNRAS* **482**, 2, 2088 (2019)

Arévalo, P., Uttley, P., *MNRAS* **367**, 2, 801 (2006)

Casella, P., Belloni, T., Homan, J., Stella, L., *A&A* **426**, 587 (2004)

Charisi, M., et al., *MNRAS* **463**, 2, 2145 (2016)

Edelson, R., Nandra, K., *ApJ* **514**, 2, 682 (1999)

Farris, B. D., Duffell, P., MacFadyen, A. I., Haiman, Z., *ApJ* **783**, 2, 134 (2014)

Gierliński, M., Middleton, M., Ward, M., Done, C., *Nature* **455**, 7211, 369 (2008)

Graham, M. J., et al., *Nature* **518**, 7537, 74 (2015)

Ingram, A., Done, C., *MNRAS* **419**, 3, 2369 (2012)

Kozłowski, S., et al., *ApJ* **708**, 2, 927 (2010)

Lehto, H. J., Valtonen, M. J., *ApJ* **460**, 207 (1996)

Li, H. Z., et al., *Publications of the Astronomical Society of the Pacific* **121**, 885, 1172 (2009)

Liu, T., et al., *The Astrophysical Journal* **803**, 2, L16 (2015)

Markowitz, A., et al., *ApJ* **593**, 1, 96 (2003)

McHardy, I. M., et al., *Nature* **444**, 7120, 730 (2006)

Merloni, A., Heinz, S., di Matteo, T., *MNRAS* **345**, 4, 1057 (2003)

Motta, S. E., et al., *MNRAS* **447**, 2, 2059 (2015)

Perez, C. A., Silbergleit, A. S., Wagoner, R. V., Lehr, D. E., *ApJ* **476**, 2, 589 (1997)

Sandrinelli, A., Covino, S., Dotti, M., Treves, A., *AJ* **151**, 3, 54 (2016)

Schwarzenberg-Czerny, A., *ApJ* **489**, 2, 941 (1997)

Stellingwerf, R. F., *ApJ* **224**, 953 (1978)

Uttley, P., McHardy, I. M., Papadakis, I. E., *MNRAS* **332**, 1, 231 (2002)

Valtonen, M. J., et al., *ApJ* **643**, 1, L9 (2006)

Vaughan, S., Uttley, P., *MNRAS* **362**, 1, 235 (2005)

Vaughan, S., et al., *MNRAS* **461**, 3, 3145 (2016)

Villata, M., Raiteri, C. M., *A&A* **347**, 30 (1999)

Wijnands, R., Homan, J., van der Klis, M., *ApJ* **526**, 1, L33 (1999)