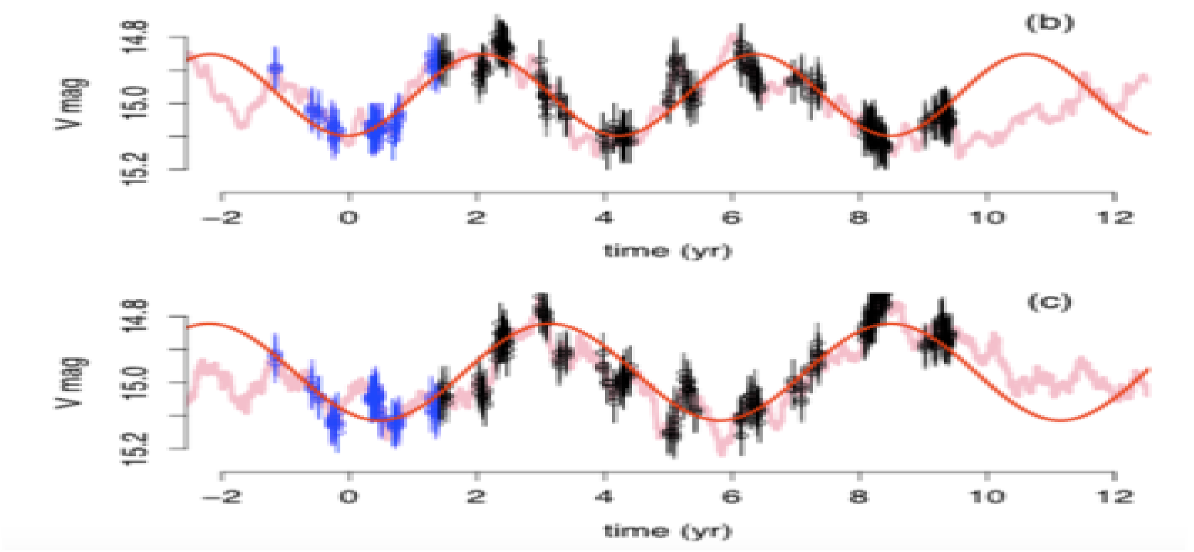
Tools for Period Searching in AGN in the Era of "Big Data"

Saikruba Krishnan[1], Alex Markowitz[1,2], A. Schwarzenberg-Czerny[1], Phil Uttley[3]

[1]Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, Warsaw, Poland
[2]UC San Diego, Center for Astrophysics and Space Sciences, La Jolla, USA
[3]Anton Pannekoek Institute of Astronomy, University of Amsterdam, Amsterdam, Netherlands

Overview

- 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.
- The community has been claiming multiple periods from various data trawls (e.g., B. Rani et al. 2009, ApJ, 696, 2170; M.J. Graham et al. 2015 MNRAS, 453, 1562; T. Liu et al. 2015, ApJ, 803, L16), spanning the full EM spectrum, and with signal periods spanning ksec to years. However, no physically-consistent picture based on these periods has emerged. We surmise that most of these periods are spurious and are due to ordinary red noise, which can mimic few-cycle sinusoid-like variability e.g, Figure~1 (from Vaughan et al. 2016 MNRAS, 461, 3145) and/or associated with improperly calibrated false alarm probabilities.

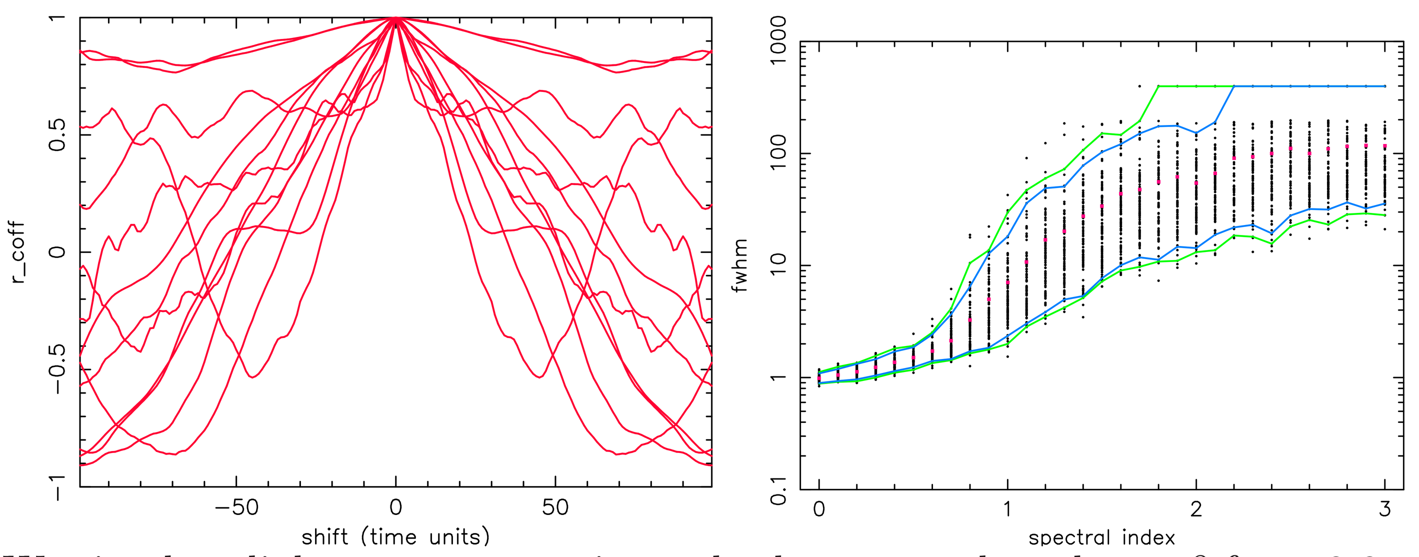


(Figure 1)

- Active Galactic Nuclei (AGN) and Black Hole X-Ray Binaries (BHXRBS) offer test-beds for General Relativity via the behavior of their dynamical accretion flows in the regime of extreme gravity. BHXRBs commonly show narrow-band, quasi-perodic oscillations (QPO) in their X-ray power spectral density (PSD) functions. QPOs originate in the inner accretion disk which directly feeds the black hole, and a range of interpretations indicate their service as tracers of black hole mass and spin, e.g., various global accretion disk oscillation modes and/or Lense-Thirring precession. Periodicities in some radio-loud AGN are interpreted as evidence of jet precession.
- The situation is urgent because we have already entered the era of "Big Data": ground-based observing programmes such as PanSTARRS, ZTF, and LOFAR now monitor large fractions of the sky, producing light curve databases which allow period searches over 10^3 to 10^6 AGN simultaneously; additional large-area monitoring programmes such as LSST and SKA will be online within the next decade.
- **Our goal is to provide guidelines on the proper use of the statistical methods which can robustly distinguish between pure stochastic red noise (no QPOs) and a mixture of red noise plus a QPO signal**. In many cases, AGN data quality is insufficient to use the Power Spectral Density function, hence to determine the underlying QPO signals we will test some of the commonly-used statistical methods: wavelets, time-domain Bayesian fitting (CARMA), epoch folding, and autocorrelation functions.
- For each tool, we will perform Monte Carlo simulations and empirically test a range of red noise continuum "background" shapes and a range of QPO strengths. We will also test a range of sampling patterns to assess the impact of data gaps inherent in realistic light curve sampling. For large surveys, we can create mock data sets assuming certain distributions in AGN mass and luminosity, and thus PSD red noise parameters; we can then gauge false alarm probabilities across entire samples. We will discuss the limitations and the range in detection sensitivity between the various methods for the various forms of the red noise background and (deterministic) characteristic signals in AGN light curves when red noise is present.

Preliminary test on the ACF

When we use the Autocorrelation Function (ACF) for searching QPO signals in AGN data, it is important to note that the pure red-noise processes (especially those with steep PSD power-law slopes) causes spurious bumps and wiggles in the ACF that some occasionally interpret as a deterministic signal (e.g., Z-Y. Zheng, N. Butler, et al. 2016, ApJ, 827, 56). So it is really necessary to understand the extent to which the ACF can or cannot be used as a substitute for the power spectrum in the limit for poor data sampling.



spectral index (β)	FWHM/Duration	Standard Deviation/Duration
0.0	0.005	3.15×10^{-4}
0.3	0.005	5.35×10^{-4}
0.6	0.008	3.90×10^{-3}
1.0	0.037	3.28×10^{-2}
1.2	0.081	0.17
1.6	0.231	0.347
2.0	0.358	0.726
2.4	0.485	0.776
2.8	0.573	0.782

Figure 2: We simulate light curves assuming unbroken power-law slopes β from 0.0 to –3.0 and plot FWHM of the central ACF peak against β (solid green line represents 98% confidence limit and the solid blue line represents 95% confidence limit). $FWHM = 400s$ denotes the ACF correlations of the light curves for which there is no FWHM obtained.

As an initial test we ask if the ACF can be used as a proxy for PSD power-law slope when there are insufficient data for PSD measurement. We use Monte Carlo simulations of unbroken power-law PSD models to test if the FWHM of the central peak is a reliable estimator for power-law slope β. We find that only for very flat slopes ($\beta \lesssim 1$), FWHM is a reasonable estimator for β; for steeper slopes, the standard deviation of FWHM is extreme.

Work in progress: We will test strengths of QPO signatures in the presence of red noise, and will adopt realistic sampling patterns (e.g., that of LSST), using the zDCF (T. Alexander 2013, arXiv:1302.1508), as it is suitable for highly uneven sampling.

Work plans on Wavelets

Wavelets are used as an tool for searching for periodic signals, with applications found across physics and mathematics, e.g., Wavelet transforms and their applications to turbulents (Farge 1992), climate science (Torrence & Compo 1998).

However, in many cases, one searches for signals relative to a white noise/poissonian background. Application to different red noise backgrounds (as appropriate for AGN) is not yet thoroughly explored.

Our plan is to map out how to use the wavelet tool to quantify (narrow-band) strictly- and quasi-periodic signals, including transient and time-variable signals, in the presence of (broadband) red noise. For example, transient QPOs have been observed in BHXRBs (e.g., Nespoli et al. 2003, A&A, 412, 235; Casella et al. 2004, A&A, 426, 587), and given similar accretion characteristics, we may expect that they exist in Seyfert AGN (e.g., M. Gierlinski et al. 2008 Nature, 455, 369; RE J1034+396).

We are in the process of formulating a methodology to translate such data/model residuals into GLOBAL confidence contours for the null hypothesis model, while accounting for the "look elsewhere effect" across both frequency and wavelet scale against which candidate outlying features can be compared.

We will test a range of PSD scenarios, and test the weighted Z-transform weighted wavelet (Foster, G. 1996, AJ, 112, 1709) to account for uneven sampling, as has been used on AGN (e.g., Wang et al. 2014, MNRAS 443, 58; Bhatta et al. 2016, ApJ, 832, 47).

These tests will allow the community to revisit previous claims of wavelets in AGN and properly re-evaluate their statistical signficances.

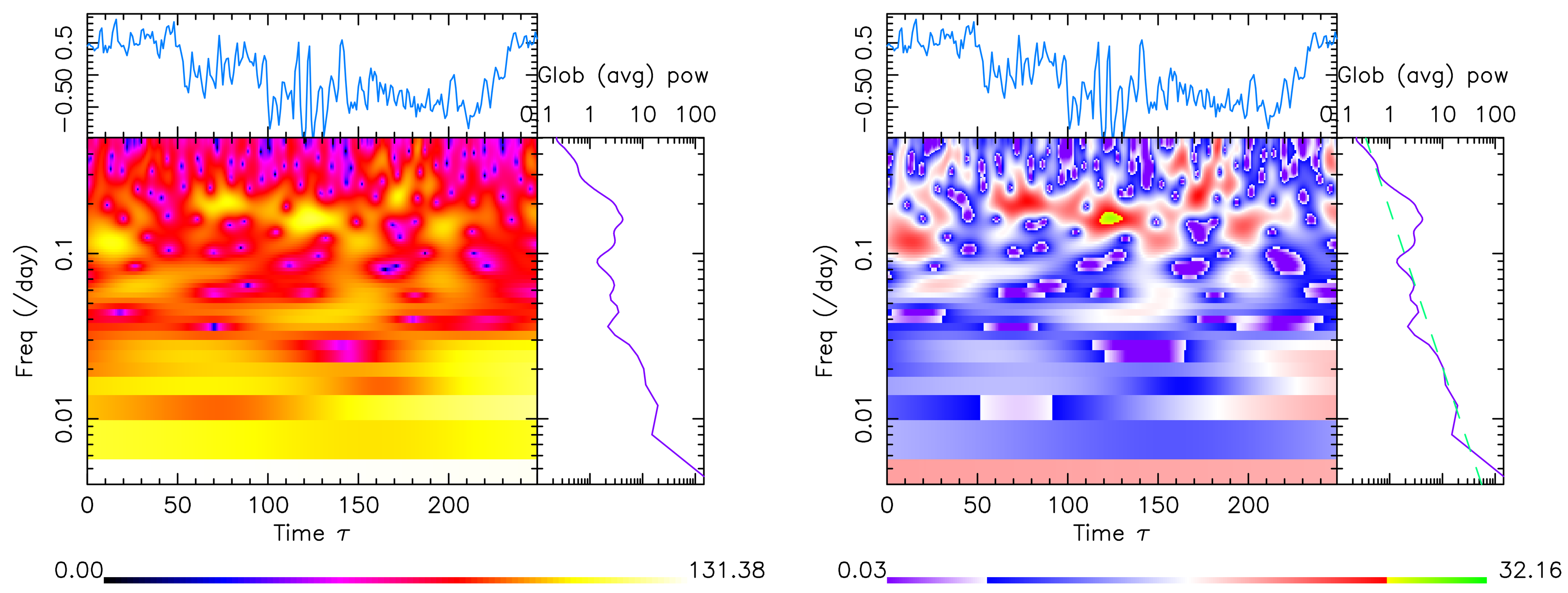


Figure 3: We assume PSD model of an unbroken power law ($\beta = 1.5$) plus a transient QPO with quality factor 4, centroid frequency 2×10^{-6} Hz, and RMS factor R increasing linearly from start/end of observation of value 0.03 to 0.57 at the middle of the observation. Thus, the ratio of variability amplitudes $F_{var}(QPO)/F_{var}(PL)$ peaked at 3.3 at the middle of the observation and ~ 0.01 at the beginning/end.
The left plot shows the WWZ powers and global periodogram; the amplitudes of power associated with the underlying red noise towards longer time scales makes it difficult to discern the QPO's presence (middle of observation, near $(5.8 \text{ days})^{-1}$. In the right panel, we perform least-squares fit to the global periodogram (dashed green line) and normalize the powers to the best fit. The ratio of the residuals are plotted on the right, revealing the excess power at 0.17 cyc/day.

Further Directions/Acknowledgements

We will adopt the algorithm of Emmanoulopoulos et al. (2013) for more complex (i.e., log-linear) flux distributions of light curve. Our results will apply to both individual light curves and database trawls, and will help the community produce only statistically-significant period claims in the literature.

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