



Variable classification in the LSST era: exploring a model for quasi-periodic light curves

J. C. Zinn,¹★ C. S. Kochanek,^{1,2} S. Kozłowski,³ A. Udalski,³ M. K. Szymański,³ I. Soszyński,³ Ł. Wyrzykowski,³ K. Ulaczyk,^{3,4} R. Poleski,^{1,3} P. Pietrukowicz,³ J. Skowron,³ P. Mróz³ and M. Pawlak³

¹Department of Astronomy, The Ohio State University, 140 West 18th Avenue, Columbus, OH 43210, USA

²Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 W. Woodruff Avenue, Columbus, OH 43210, USA

³Warsaw University Observatory, Al. Ujazdowskie 4, PL-00-478 Warszawa, Poland

⁴Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

Accepted 2017 March 6. Received 2017 March 6; in original form 2016 December 14

ABSTRACT

The Large Synoptic Survey Telescope (LSST) is expected to yield $\sim 10^7$ light curves over the course of its mission, which will require a concerted effort in automated classification. Stochastic processes provide one means of quantitatively describing variability with the potential advantage over simple light-curve statistics that the parameters may be physically meaningful. Here, we survey a large sample of periodic, quasi-periodic and stochastic Optical Gravitational Lensing Experiment-III variables using the damped random walk (DRW; CARMA(1,0)) and quasi-periodic oscillation (QPO; CARMA(2,1)) stochastic process models. The QPO model is described by an amplitude, a period and a coherence time-scale, while the DRW has only an amplitude and a time-scale. We find that the periodic and quasi-periodic stellar variables are generally better described by a QPO than a DRW, while quasars are better described by the DRW model. There are ambiguities in interpreting the QPO coherence time due to non-sinusoidal light-curve shapes, signal-to-noise ratio, error mischaracterizations and cadence. Higher order implementations of the QPO model that better capture light-curve shapes are necessary for the coherence time to have its implied physical meaning. Independent of physical meaning, the extra parameter of the QPO model successfully distinguishes most of the classes of periodic and quasi-periodic variables we consider.

Key words: methods: data analysis – stars: oscillations – stars: variables: Cepheids – stars: variables: general – stars: variables: RR Lyrae – Magellanic Clouds.

1 INTRODUCTION

Time domain astronomy has resulted in a range of observed astrophysical variability regimes that may be broadly categorized as periodic, quasi-periodic, stochastic and transient. Transient events are short time-scale changes in flux such as gamma-ray bursts (Klebesadel, Strong & Olson 1973; Cano et al. 2016), supernovae (SNe; for a review see e.g. Woosley & Weaver 1995), cataclysmic variables (Robinson 1976) and stellar flares (Variable Star Network; Kato et al. 2004). These can frequently be modelled with a template light curve whose structure and parameters can be used to classify the transient (e.g. photometric SN classification; Pskovskii 1977; Hamuy et al. 1996; Sako et al. 2008, 2011). Periodic variables such as RR Lyrae, Cepheids and eclipsing binaries are classified based

on their period and the structure of their phased light curves (e.g. Pojmański 2002; Debosscher et al. 2007; Soszyński et al. 2008, 2009a; Sarro et al. 2009; Graczyk et al. 2011). Quasi-periodic sources such as Miras or spotted stars have one or more dominant frequencies describing their variability, but do not maintain consistent phase and/or amplitude as a function of time (e.g. Howarth 1991; Bedding et al. 2005). In this regime, sources are classified by a period and average light-curve structure (usually Fourier component ratios; e.g. Soszyński et al. 2008), but there is no measure of the coherence of the variability. Stochastic variables such as active galactic nuclei (AGN) vary without any obvious pattern and may be modelled by stochastic processes (e.g. Kelly, Bechtold & Siemiginowska 2009; Kozłowski et al. 2010; MacLeod et al. 2010; Zu et al. 2016).

While studies of periodic sources are ubiquitous in astronomy, there are far fewer quantitative studies of quasi-periodic systems. Presently, the primary large-scale application is the determination

* E-mail: zinn.44@osu.edu

of stellar rotation rates for *Kepler* stars (McQuillan, Aigrain & Mazeh 2013; Reinhold, Reiners & Basri 2013; Balaji et al. 2015). In this case, the variability is quasi-periodic due to the finite lifetime of star spots and differential rotation. These periods have been used to study rotation rates in stars (e.g. Frasca et al. 2011) and are the basis for gyrochronology, where stellar ages are estimated based on models for spin-down rates (García et al. 2014). Quasi-periodic variables such as long-period variables (LPVs) are also well known. Qualitative measures of irregularity in light curves have been used for classification, most notably in the General Catalogue of Variable Stars (Kholopov et al. 1985) and the Optical Gravitational Lensing Experiment Catalogue of Variable Stars (Soszyński et al. 2008), but quantification of the degree of quasi-periodicity for these sources remains little explored.

The study of variable sources benefits from a broad range of surveys. There are many ongoing surveys for transients such as the Optical Gravitational Lensing Experiment (OGLE; e.g. Udalski et al. 2008), the Catalina Real-Time Transient Survey (Drake et al. 2009), the Palomar Transient Factory (Law et al. 2009), QUEST (Snyder 1998), the All-Sky Automated Survey (Pojmański 1997) and the All-Sky Automated Survey for SuperNovae (Shappee et al. 2014). Apart from these variability surveys, searches for transiting exoplanets such as the Hungarian Automated Telescope (HAT; Bakos et al. 2002), *Kepler* (Borucki et al. 2010), CoRoT (Auvergne et al. 2009), KELT (Pepper, Gould & Depoy 2004) and SuperWASP (Pollacco et al. 2006; Smith et al. 2014) have also produced a wealth of light curves for other variable phenomena. More general surveys such as the Sloan Digital Sky Survey (SDSS; York et al. 2000), Pan-STARRS (Kaiser et al. 2002) and the Dark Energy Survey (The Dark Energy Survey Collaboration 2005) also include time domain components. Future surveys such as the Large Synoptic Survey Telescope (LSST; Tyson 2002; Ivezić et al. 2008; LSST Science Collaboration et al. 2009), TESS (Ricker et al. 2014), PLATO 2.0 and the Zwicky Transient Facility (Smith et al. 2014) promise to expand time domain studies still further.

Two challenges to science with variability surveys are to first classify the variable sources and then to extract physical information about the variability from the light curves. A promising avenue for parsing the huge numbers of light curves is machine learning (e.g. Woźniak et al. 2004; Debosscher et al. 2007; Bloom et al. 2012; Richards et al. 2012; see Ball & Brunner 2010; Kim et al. 2014 for reviews of machine learning in astronomy). In most machine learning approaches, a large number of statistics are calculated for each light curve and an algorithm is trained to classify variables based on the properties of a sample of previously classified variables (but see e.g. Mackenzie, Pichara & Protopapas 2016, for an example of unsupervised classification of variable stars). While these methods work well for identifying transients such as SNe and highly periodic sources like Cepheids and RR Lyrae, stars with irregular periodicity and multiperiodicity have proven more difficult to accurately classify. One problem may be that the statistics employed in many of the current approaches have no physical significance. Parameters with physical meaning such as periods may be better for classifying light curves than ad hoc statistical measures, in addition to providing insights into the nature of the variable source. The challenge is the definition and extraction of new physically meaningful parameters from quasi-periodic or stochastic light curves.

The introduction of stochastic process models for quasar variability (Kelly et al. 2009; Kozłowski et al. 2010) is a good recent demonstration of these principles. A damped random walk (DRW) stochastic process characterized by an amplitude and a coherence time-scale provides a good, compact statistical model of quasar

variables on time-scales of days to decades (MacLeod et al. 2012). The parameters of the model are then correlated with the observed wavelength and the intrinsic properties of the quasars (e.g. MacLeod et al. 2010), potentially providing insight into the physical origin of the variability (Kelly et al. 2009; Dexter & Agol 2011; Andrae, Kim & Bailer-Jones 2013). There is evidence for deviations from the DRW, particularly on short time-scales (e.g. Mushotzky et al. 2011; Andrae et al. 2013; Zu et al. 2013; Graham et al. 2014; Kasliwal, Vogeley & Richards 2015), and this may allow the extraction of additional physical scales from AGN light curves. The DRW model has also seen application in He et al. (2016) to model Miras, in which they find that adding a DRW component in addition to a purely periodic component can improve the fidelity of period estimates by up to 20 per cent.

The DRW model is the first ($p = 1$) in a series of continuous time autoregressive moving average (CARMA(p, r)) processes, which are the solutions to stochastic linear differential equations of order p driven by white noise and up to $r < p$ derivatives of white noise processes. The higher order processes in this sequence have occasionally been used in astronomy (e.g. Koen 2005, 2012; Andrae et al. 2013; Kelly et al. 2014; Kasliwal et al. 2015; Aigrain, Parviainen & Pope 2016). In particular, Kelly et al. (2014) offer a package implementing fits of these models to light curves based on methods from the forecasting literature (e.g. Brockwell & Davis 2002).

Physically, the $p = 2$ process models a quasi-periodic oscillation (QPO) characterized by an amplitude, a period and a coherence time. Koen (2005, 2012), Kelly et al. (2014) and Aigrain et al. (2016) give examples of applying the model to periodic and quasi-periodic variables, but it has never been employed for a large-scale variability survey. Here we carry out such a survey, applying the QPO model to large numbers of variable stars and quasars in the OGLE-III fields for the Large Magellanic Cloud (LMC). We also compare the QPO results to those from modelling the same sources using a DRW, and explore to what extent certain variables prefer a QPO to traditional purely periodic models. The goals are to explore whether the two models successfully distinguish between variable stars and quasars and whether the QPO coherence time provides a useful means of classifying and understanding variable stars. In Section 2, we describe the light-curve data and the model formalism. In Section 3, we provide a guide towards understanding the results presented in Section 4, and we conclude in Section 5 with a discussion.

2 DATA AND METHODS

The OGLE observes the Magellanic Clouds, and portions of the Galactic bulge and disc in the I and V bands using the 1.3-m Warsaw Telescope at the Las Campanas Observatory. We use light curves of sources in or in front of the LMC from the OGLE-III survey (Udalski et al. 2008), which operated from 2001 to 2009. We examine 3361 classical Cepheids (Soszyński et al. 2008), a subset of 50 000 LPVs (Soszyński et al. 2009b) and 24 906 RR Lyrae (Soszyński et al. 2009a), as well as 753 quasi-stellar objects (QSOs) identified in the Magellanic Quasars Survey (MQS; Kozłowski et al. 2013). We used only the I -band data as they have a better mean cadence (~ 6 d) over a total baseline of ~ 3000 d. We removed MQS QSOs fainter than $I = 19.5$ mag from our analysis due to degraded light-curve quality. We require that each object strongly prefer (at 95 per cent confidence) either a DRW and/or a QPO model above pure white noise, leaving 1460 Cepheids, 24 542 LPVs, 24 223 RR Lyrae and 443 QSOs (details below). This requirement ensures

that sufficient correlations can be detected in the light curve to distinguish a stochastic process model from a simple broadening of the photometric uncertainties (a ‘white noise’ model). The errors on the OGLE-III data points are calculated using the difference image analysis (DIA) software described in Woźniak (2000). In many cases, these errors may be underestimated. The errors on the QSOs and Cepheids can be corrected by rescaling the DIA errors to $\sigma'_i = \sqrt{(\gamma\sigma_i)^2 + \epsilon^2}$, to match the observed dispersion in standard stars as a function of magnitude, where γ and ϵ are computed on a field-to-field basis as described in Wyrzykowski et al. (2009).

The OGLE LPVs have been broadly categorized based on light-curve properties into semiregular variables (SRVs), Miras and OGLE small-amplitude semiregular variables (OSARGs) by Soszyński et al. (2009b). All three types are understood to be giant branch or asymptotic giant branch stars. OSARGs have variability amplitudes ranging from 0.005 to 0.13 mag with periods ranging from 10 to 100 d (Wray, Eyer & Paczyński 2004), and the pulsation mechanisms responsible for their light curves are not well understood (but see Christensen-Dalsgaard, Kjeldsen & Mattei 2001; Soszyński et al. 2004). Miras are generally believed to be variable as a result of inherently unstable radial modes due to opacity effects in the atmosphere (e.g. Wood 1974), with periods ranging from 100 to 1000 d and amplitudes of up to 10 mag. SRVs have similar periods to Miras, but are distinguished from Miras by their lower amplitudes and because they pulsate in both the fundamental and the overtone radial modes, instead of only the fundamental mode. Although their light curves exhibit qualitatively different light curves, with SRVs generally having less coherent periodicity than Miras, Soszyński, Wood & Udalski (2013) hypothesize that as red giants evolve, they first pulsate as OSARGs, then as SRVs and finally as Miras (see also Bedding & Zijlstra 1998).

Among the best-studied variable stars are Cepheid variables – giant stars with pulsation periods mostly from 2 to 20 d and amplitudes from 0.2 to 1 mag in V band due to the κ -mechanism (Zhevakin 1953). The OGLE Cepheids are divided into subclasses based on their pulsation modes, from the fundamental mode (‘F’) and first overtone (‘1’) pulsators up to the third overtone pulsators, with a small number of multimode pulsators (Soszyński et al. 2008). RR Lyrae stars are horizontal branch stars in the instability strip that also pulsate due to the κ -mechanism (Christy 1966). Based on their periods and light-curve shapes, RR Lyrae are subdivided into RRab (fundamental mode pulsators), RRc (first overtone pulsators), RR_e (second overtone pulsators) and RR_d (simultaneous fundamental and first overtone pulsators). Typical variability amplitudes for RR Lyrae range from 0.2 to 2 mag in V band, and the pulsation periods range from 0.2 to 1 d.

We model light curves using a family of stochastic process models known as continuous-time autoregressive moving averages (CARMA). These models statistically describe the solution to stochastically driven ordinary differential equations (ODEs) by way of an autocovariance function, $S \equiv \langle ss^T \rangle$, whose structure depends on the order of the ODE, p and the weights given to the $r < p$ noise processes driving the system. For the purposes of this paper, we call the $p = 1$ case a DRW and the $p = 2$ case a QPO. The autocovariance functions of the CARMA processes are the sum of p exponentials $e^{\omega_i t}$, where the ω_i must be real or appear in complex conjugate pairs (see Brockwell & Davis 1996). The DRW model is the lowest order example, and its covariance function between times t_i and t_j is

$$S_{ij} = \sigma_{\text{DRW}}^2 \exp(-|t_j - t_i|/\tau), \quad (1)$$

where σ_{DRW} describes the variance of the light curve on long time-scales and τ is the coherence time. We use the DRW for comparison to the QPO model in order to explore whether a quasi-periodic model is a better model for the stellar variable sources (or the reverse for quasars). The QPO model is the case $p = 2$, with a complex conjugate pair of w_i , which we can divide into an oscillation period, $P = 2\pi/\omega$, and a coherence time, τ , to give the autocovariance function

$$S_{ij} = \frac{\sigma^2}{2} \exp[-|t_j - t_i|/\tau] \cos[\omega(t_j - t_i)]. \quad (2)$$

We could also add a phase, but we omit a phase factor in favour of assuming that all real variable sources have a maximum in their autocovariance functions as $|t_j - t_i| \rightarrow 0$.

Whereas the DRW τ parameter indicates the time-scale over which variability is correlated, the QPO τ parameter is the time-scale over which the process appears as a coherent oscillation. Thus, the ratio of the coherence time to the period, τ/P , roughly corresponds to the number of oscillations over which the light curve will maintain phase coherence. Fig. 1 illustrates this meaning of the coherence time with a series of phased light curves ranging from an essentially periodic Cepheid at the top to a non-periodic MQS quasar at the bottom. We adopt $\tau/P > 10^4$ as our operational definition of periodic, although RR Lyrae and Cepheid variables can be even more strictly periodic. In practice, we cannot distinguish among models with $\tau > 10^4$ d due to the OGLE time baseline (cf. the discussion of Fig. 2 below; see also Kozłowski 2017). When fitting models, we impose $\tau < 10^4$ d. In the limit that $\tau/P \ll 1$, illustrated by the quasar, the QPO model becomes the DRW model.

In addition to the process model for the variability, we assume the data have uncorrelated Gaussian errors (noise) with an autocovariance function $\mathbf{N} \equiv \langle nn^T \rangle = \delta_{ij}\sigma_i^2$. The data are then described by the total covariance matrix, $\mathbf{C} = \mathbf{S} + \mathbf{N}$, which combines the covariance due to the signal S and the covariance due to the noise N . The model also includes a set of linear parameters to deal with the light-curve means and long time-scale trends (see Appendix). In all cases, we use the linear parameters to remove the light-curve mean. As we explore in Section 3, long-term trends can affect the results for the LPVs. We experimented with removing polynomial trends up to third order, ultimately settling on using a cubic trend model for all stellar sources (LPVs, RR Lyrae and Cepheids), and two trends for QSOs corresponding to the OGLE-III and OGLE-IV portions of the light curves because they can be offset. We only used the OGLE-III light curves for the variable stars so there was no need to allow for mean offset other than for the quasars. Our method automatically includes the uncertainties in the linear parameters into the uncertainties in the process parameters.

The final likelihood function in this framework is then

$$\mathcal{L} = \frac{\exp(-(y - \mathbf{L}\hat{q})^T \mathbf{C}^{-1}(y - \mathbf{L}\hat{q})/2)}{(2\pi)^{P/2} |\mathbf{C}|^{1/2} |\mathbf{L}^T \mathbf{C}^{-1} \mathbf{L}|^{1/2}}, \quad (3)$$

where $\mathbf{L}\hat{q}$ is the best-fitting trend model described by the matrix, \mathbf{L} , and the best-fitting linear coefficients, \hat{q} (see Appendix), y is the observed light curve and P is the number of parameters in the stochastic process model. The computational barrier to performing such an analysis lies in the inversion of the covariance matrix, \mathbf{C} . Rybicki & Press (1992) found a method to reduce the scaling from $\mathcal{O}(N^3)$ to $\mathcal{O}(N)$ for the cases $p = 1$ and $p = 2$, and this was the method used by Kozłowski et al. (2010) and MacLeod et al. (2010) to rapidly analyse OGLE-III and SDSS light curves with the DRW model. Ambikasaran et al. (2014) have subsequently found an algorithm that is $\mathcal{O}(Np^2)$, for all p , and we

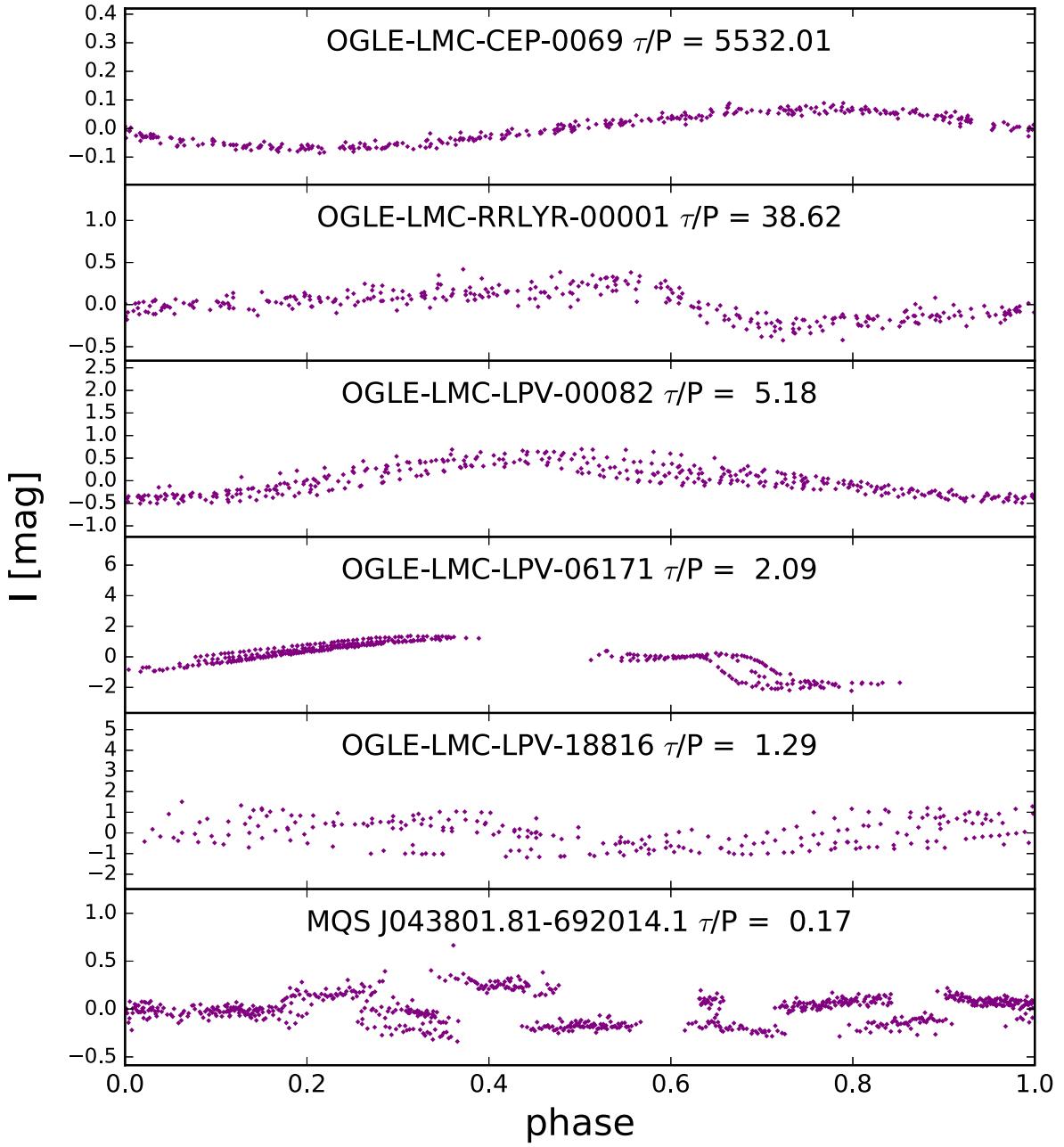


Figure 1. Folded light curves illustrating the meaning of the coherence time-scale in the QPO model. Each panel shows a folded light curve of a source ranging from coherent ($\tau/P \gg 1$) sources at the top to incoherent ($\tau/P \ll 1$) sources at the bottom. The most coherent source shown is the Cepheid OGLE-LMC-CEP-0069, followed by an RR Lyrae, a series of LPVs and ending with the quasar MQS J043801.81–692014.1. That τ/P is not larger for OGLE-LMC-RRLYR-00001 is discussed in Section 3.

use this sparse matrix method. The forecasting algorithms used by Kelly et al. (2009, 2014) are also $\mathcal{O}(N)$, but in Kozłowski et al. (2010) we found that evaluating the full likelihood given by equation (3) had significantly more statistical power than the forecasting methods.

We model the light curves by first finding best-fitting parameters that maximize the likelihood (equation 3) using a simulated annealing method, as implemented in PYGMO.¹ Our problem necessitates a maximum likelihood estimation (MLE) approach for optimization

as opposed to a grid search because the length of the OGLE light curves requires large numbers of period samples even before adding extra parameters. For a fixed phase error, $\delta\phi$, the period has to be sampled by $\Delta P/P \sim (\delta\phi/2\pi)(P/T)$, where $T \sim 3000$ d is the baseline of OGLE light curves. For a short-period Cepheid (RR Lyrae) period $P \sim 1$ d ($P \sim 0.1$ d) and $\delta\phi \sim 1$, this requires $\Delta P/P \sim 1/3000$ ($1/30\,000$). Using the MLE estimate for the best-fitting DRW or QPO model as a starting point, we then sample the parameter space using the PYTHON Markov chain Monte Carlo (MCMC) package EMCEE (Foreman-Mackey et al. 2013). We use standard logarithmic priors for the process parameters. Our chain length is significantly larger than the autocorrelation time (e.g. Sokal 1996) of the

¹ <https://github.com/esa/pagmo/tree/master/PyGMO>

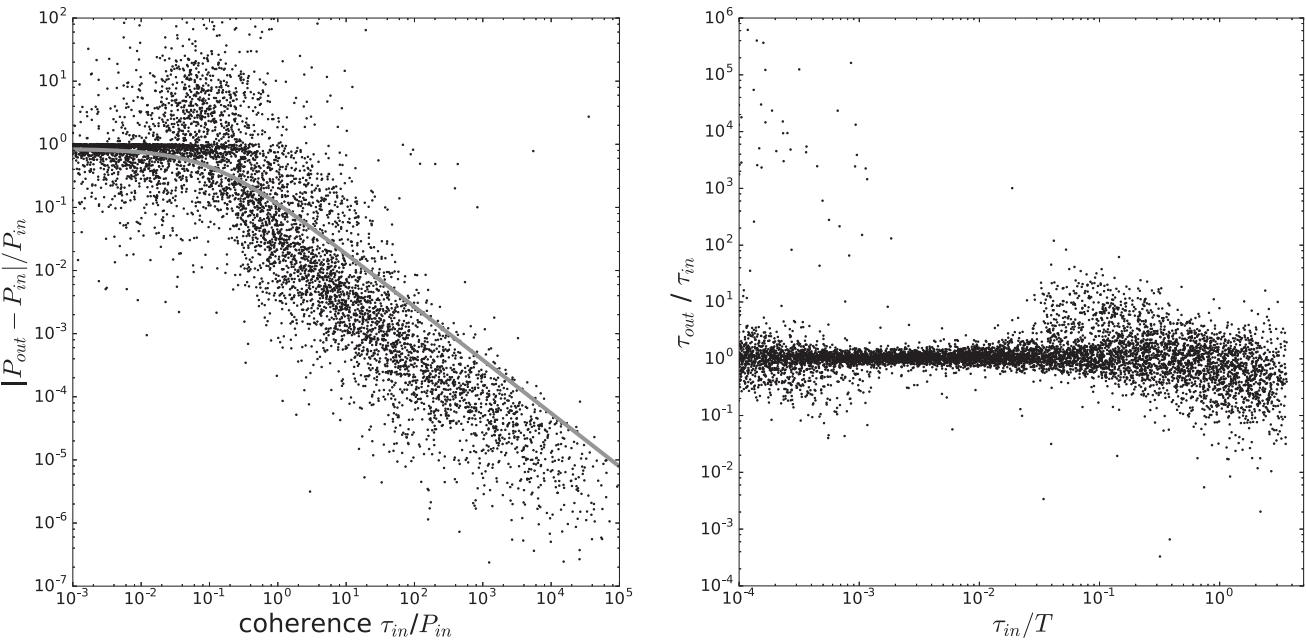


Figure 2. Fractional error in period, $|P_{\text{out}} - P_{\text{in}}|/P_{\text{in}}$ as a function of the coherence of the oscillation, $\tau_{\text{in}}/P_{\text{in}}$, for artificial light curves (left). A simple model for the fractional error in period as a function of the coherence is shown as a solid line to guide the eye. Recovered versus input coherence time, $\tau_{\text{out}}/\tau_{\text{in}}$, as a function of the coherence time in units of the light-curve baseline, τ_{in}/T , where $T \approx 3000$ d (right). Coherence times approaching the length of the light-curve time baseline or shorter than the cadence ($\tau_{\text{in}}/T = 1 \times 10^{-3}$) are increasingly difficult to recover.

estimated parameters calculated by EMCEE for most light curves. In this limit, the mean marginalized parameter estimated is representative of the true mean of the parameter. Parameter errors are quoted at 1σ , and the best-fitting parameter is taken to be the mean of the MCMC chains.

To compare models with differing numbers of parameters, we use the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). For a light curve with likelihood \mathcal{L} , N data points, and k model parameters, the Bayesian likelihoods (equation 3) are modified to be

$$\text{AIC} \equiv -2 \ln \mathcal{L} + 2k \quad (4)$$

and

$$\text{BIC} \equiv -2 \ln \mathcal{L} + k \ln N. \quad (5)$$

The BIC penalizes the addition of new parameters more heavily than does the AIC. These allow us to compare models with differing numbers of parameters, and, in particular, address the questions of which of the models (DRW or QPO) better fits the data, and whether a quasi-periodic model is a better fit than a periodic model. We present the AIC results. Our initial criterion for rejecting sources better fit by white noise (the limit $\tau \rightarrow 0$ d for either QPO or DRW) corresponds to rejecting objects with $\Delta \text{AIC} < -1$ between the white noise and the DRW/QPO models. In order to compress the full range of the likelihood differences between models 1 and 2, we show the distributions in terms of

$$f(x) = \sinh^{-1} \left[\frac{\text{AIC}_1 - \text{AIC}_2}{2} \right] = \sinh^{-1} x. \quad (6)$$

This transformation provides a linear scaling for $|x| \lesssim 1$ and transitions to a logarithmic scaling $\ln|x|$ for $|x| \gg 1$. The normalization is chosen such that the likelihood ratio is 95 per cent at $f = \pm 1$ (corresponding to $\Delta \chi^2 = 4$ for χ^2 statistics). For $f < -1$ ($f > 1$),

the source is best described by model 1 (model 2), and at $|f| < 1$, the maximum likelihood probability for model 1 (model 2) is greater than 5 per cent. The probability of the less favoured model plummets outside the $|f| < 1$ region, falling below 0.07 per cent for $|f| > 2$.

3 UNDERSTANDING THE RESULTS

We next discuss a series of experiments to help understand the global results. We start by simply confirming that the process is internally consistent by generating QPO model light curves with cadences, amplitudes and uncertainties typical of the OGLE data and showing that we can recover the input parameters reasonably well. We also show that we can recover the periods of the LPVs with some dependence on the treatment of long time-scale trends in the light curves. For shorter period variables, there is essentially never a problem in estimating the period. When we then applied the method to the actual light curves, we found that the results showed a significant dependence on the light-curve shape, behaving as expected for relatively sinusoidal light curves but with counterintuitive results for very non-sinusoidal variables like long-period Cepheids. In particular, it was surprising to find that QPO models of extremely periodic RRab and long-period Cepheids frequently preferred low-coherence QPO models or even the DRW model. This motivated a series of experiments modelling purely periodic sine and triangle waves as a function of the noise relative to the amplitude that demonstrate the effect of light-curve shape on QPO parameter estimation. Finally, we introduce a higher order, $p = 4$, QPO model to confirm that a higher order model that can better mimic the structure of less sinusoidal light curves provides a means of addressing the problem. To provide a guide to interpreting the results in Section 4, we first tested how well we can recover QPO parameters simply due to the modelling process itself. We generated a sample of 12 000 artificial QPO light curves using cadences typical of the OGLE light curves

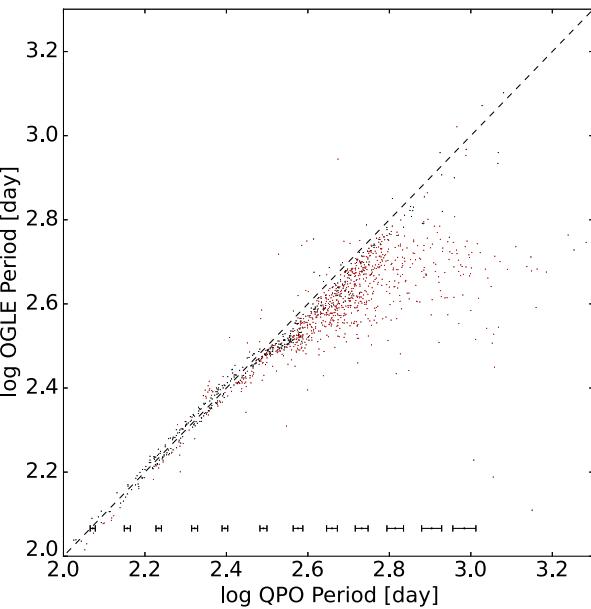
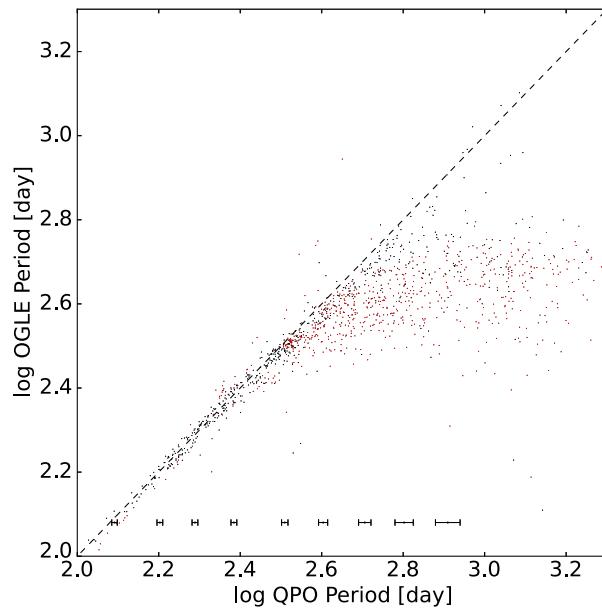


Figure 3. A comparison of QPO and OGLE period estimates for Miras using either cubic (left) or constant (right) temporal trend models. The diagonal line represents a one-to-one correspondence and the error bars along the bottom indicate the median QPO period uncertainties for a series of logarithmic bins in period. The red points in each panel are the 3σ period outliers in the other panel. In general the cubic trend model provides better agreement and is adopted as our standard model.



and with amplitudes, periods and coherence time-scales covering the range of the variables we study in Section 4, with periods varying from 0.1 to 10^3 d, τ varying from 0.1 to 10^6 d and amplitudes varying from 0.01 to 1 mag. We used a fixed photometric uncertainty of 0.02 mag, which is roughly twice the 95th percentile of the photometric error distribution of the OGLE light curves. We generated the light curves using the Cholesky decomposition method outlined in Zu, Kochanek & Peterson (2011). If the covariance matrix, \mathbf{C} , can be Cholesky decomposed into $\mathbf{C} = \mathbf{M}\mathbf{M}^T$, and \mathbf{R} is a vector of Gaussian random deviates of unit amplitude, a light-curve realization is simply $\mathbf{y} = \mathbf{MR}$.

Fig. 2 illustrates the results. Fig. 2(a) shows that the fractional errors in the periods are well modelled by $|\Delta P|/P = 1/[B + (\tau/P)^C]$ with $B \simeq 14$ and $C \simeq 0.73$, as might be expected since the ‘line width’ of the oscillation must grow as the periodicity becomes less coherent. If the damping scale is short compared to the overall light-curve length, $\tau/T \lesssim 0.1$, then the median fractional uncertainty in τ is ≈ 10 per cent, with a scatter of ± 0.3 dex (see Fig. 2). The uncertainty in τ becomes large as it approaches the light-curve time span, T , generally in the sense that it can be greatly overestimated. That τ is poorly recovered for $\tau/T \gtrsim 1$ justifies our definition of a purely periodic variable as one where $\tau = 10^4$ d ($\tau \approx 10T$) is the best model. We also demonstrate in Fig. 2(b) that τ is poorly recovered for $\tau/T \lesssim 10^{-3}$, which is the typical sampling time-scale (~ 3 d). Studies of DRW models for quasars have also found that the damping time-scale is difficult to recover accurately from typically existing data where τ is comparable to T (MacLeod et al. 2010; Kozłowski 2017).

We also tested the ability of the QPO models to recover the published OGLE-III periods of the variable stars. We find that we can always recover the period of short-period variables, with periods $\lesssim 100$ d. Care must be taken, however, to adequately resolve period space when searching for periods much less than the baseline of the light curve ($T \sim 3000$ d, for OGLE light curves) because the like-

lihood peaks become increasingly narrow due to the large number of periods spanned by the data.

As shown in Fig. 3 for the Miras, the situation is quite different for the LPVs. The period agreement is very good for shorter periods ($P < 300$ d), but there are significant numbers of cases where the QPO period estimate is significantly longer than the OGLE period. It is not surprising that the discrepancies are concentrated at longer periods since even for truly periodic sources the fractional period uncertainty will be on the order of the period divided by the time span of the data. However, the discrepancies are larger than we would expect based on our simulations of modelling randomly generated QPO light curves (see Fig. 2b). Fig. 4 shows the light curves and QPO models for two examples, LPV-00144 where the period discrepancy is large ($P_{\text{OGLE}} = 364$ d versus $P_{\text{QPO}} = 577$ d) and LPV-00743 where the period discrepancy is smaller ($P_{\text{OGLE}} = 216$ d versus $P_{\text{QPO}} = 209$ d). The obvious distinction is that LPV-00743 lacks the large amplitude changes of LPV-00144. Such large-scale features are common among the period outliers.

This suggested that the period differences might in part be driven by the cubic model for long-term trends that is included in our standard analysis, so we repeated the calculations removing constant (light-curve mean), linear and quadratic trends. Fig. 3 compares the results for simply removing a constant to the default cubic model, and we see that the number of outliers increases significantly. We also show in Fig. 3 where outliers for the two trend models lie in the distribution found for the other trend model. Given the results of these experiments, we kept the cubic model as our default for stellar sources. The trend model choice is also important for the recovery of DRW parameters, particularly for QSOs. Unsurprisingly, fitting trend models of higher order than a simple mean results in removing genuine stochastic behaviour and biases the estimates of τ towards artificially small or high values (not shown). For this reason, we opted to only remove a mean each for the OGLE-III and OGLE-IV sections of the QSO light curves for the DRW model. A cubic was

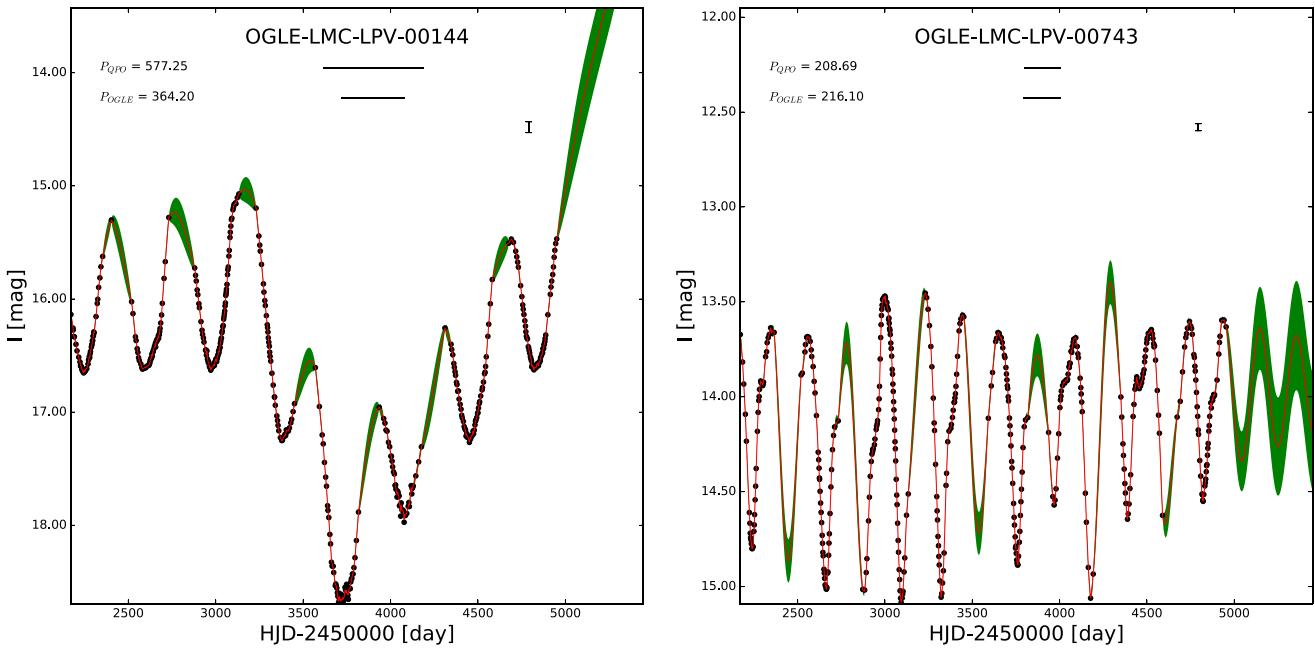


Figure 4. Miras with large (left) and small (right) period differences between the OGLE catalogue and the QPO models. The points are the OGLE data and the curve is the average of QPO models at the indicated, fixed P_{QPO} that fit the data well. The green ‘error snake’ represents the 1σ dispersion of these light-curve models around the mean. Median photometric errors are shown as a single vertical bar. The periods are listed in each panel along with lines to graphically show the scale of the periods and their differences relative to the light curves. The divergence of the light-curve model in the left-hand panel is a consequence of the cubic trend model.

removed for each of the two light-curve sections when fitting QPO models to QSOs.

One technical point to emphasize about Fig. 4 is that the QPO model curve is not a particular ‘best-fitting’ light curve, but the statistical average of light curves consistent with the data (see the discussion in Rybicki & Press 1992). The ‘error snake’ surrounding this curve is the 1σ dispersion of these light curves around their mean. A particular statistical realization of a light curve consistent with the data would track the mean with deviations statistically bounded by the ‘error snake’, and the construction of such constrained light-curve realizations is also described in Rybicki & Press (1992). These constrained realizations will show more small-scale structure than the mean light curve, and this is a large part of the variance captured by the ‘error snake’.

As we developed our analysis, we found (see Section 4) that periodic variables with light curves that strongly deviated from sinusoids (e.g. fundamental mode Cepheids and ab-type RR Lyrae) were frequently either favouring QPO models with short coherence times or even the DRW model over the coherent, periodic limit of the QPO model. Fig. 1 illustrates the issue for the RRab variable OGLE-LMC-RRLYR-00001, which is clearly a coherent periodic variable over the OGLE baseline, but has a best-fitting QPO coherence time of $\tau \sim 30$ d for a period of $P \simeq 0.64$ d so that $\tau/P \sim 40$ instead of $\tau/P \sim \infty$.

To test the hypothesis that light-curve shape could bias coherence time estimates and model preference, we modelled purely periodic sinusoid and triangle wave variability. Artificial light curves were generated using the same I -band magnitude (18.8 mag), period, cadence and photometric error $\sigma_i = 0.1$ mag as the variable OGLE-LMC-RRLYR-00001 shown in Fig. 1. We want to consider two possible effects. The first is the consequence of misunderstanding

the photometric uncertainties and the second is the consequence of the ‘signal-to-noise’ ratio of the variability. Fig. 5(a) shows the effect on the recovered τ of rescaling the reported noise by a factor of f_{err} , meaning that we analyse light curves with the error defined to be $f_{\text{err}}\sigma_i$ instead of the true error, σ_i . The light curve itself is not modified. Underestimating the true error should drive the solution towards $\tau \rightarrow 0$ d since this is the limit corresponding to photometric white noise, while overestimating the error should do the reverse. As we see in Fig. 5(a), the actual variable and the two model variables follow this basic trend. For the sinusoid, the best solution is periodic even when the uncertainties are deliberately underestimated, while the sawtooth requires slightly overestimated errors to be seen as a periodic oscillator. The actual RR Lyra approaches the periodic solution if the true uncertainties are $f_{\text{err}} = 1.5\text{--}2.0$ times the raw OGLE estimates rather than the recommended scaling by $f_{\text{err}} \approx 1.2$ (Wyrzykowski et al. 2009).

Fig. 5(b) shows model preference (right-hand axis) and the χ^2/dof (left-hand axis) as a function of f_{err} . Model preference is determined by the AIC criterion and the scaling defined in equation (6), for which a positive (negative) value indicates a preference for the periodic (QPO) model. The sine, sawtooth and OGLE-LMC-RRLYR-00001 light curves all converge to a periodic model for a sufficiently large f_{err} , essentially tracking the τ parameter estimates. The goodness of fit shows two branches. If the errors are greatly underestimated ($f_{\text{err}} \rightarrow 0$), the data can be well modelled by the white noise limit of the QPO process ($\tau \rightarrow 0$ d). If the errors are overestimated, the data can be overfit ($\chi^2/\text{dof} < 1$) and the process can be periodic ($\tau \rightarrow \infty$ d). For the sinusoid, we see that for $f_{\text{err}} = 1$, we have $\chi^2/\text{dof} = 1$ and $\tau \rightarrow 10^4$ d, as we would expect. For the sawtooth, the fit at $f_{\text{err}} = 1$ is poor ($\chi^2/\text{dof} \simeq 1.4$), the best-fitting model is only moderately coherent ($\tau \approx 170$ d) and the QPO model

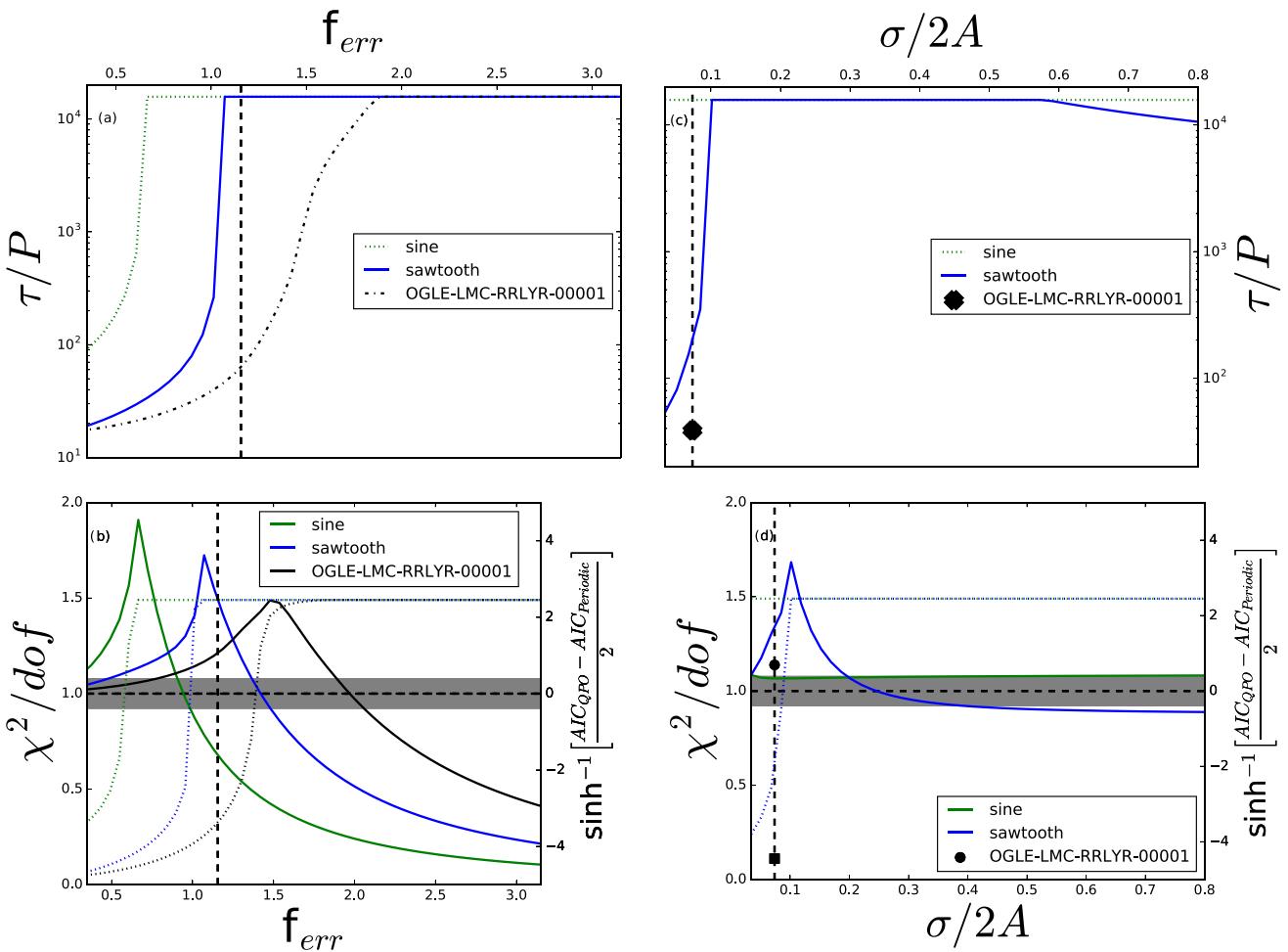


Figure 5. The upper panels show the coherence ratio, τ/P , for the OGLE RR Lyra OGLE-LMC-RRLYR-00001 and artificial light curves with sinusoidal or sawtooth waveforms. The lower panels show the QPO model preference (equation 6; dotted lines, right-hand scale) and the χ^2/dof (solid lines, left-hand scale); positive/negative values favour periodic/QPO. The vertical dashed line corresponds to the recommended factor by which the reported OGLE-III errors for OGLE-LMC-RRLYR-00001 should be increased (Wyrzykowski et al. 2009). The horizontal dashed line corresponds to a χ^2/dof of unity and the grey band indicates the expected 68 per cent confidence region assuming Gaussian statistics. The left-hand columns show these quantities as a function of the error scaling factor f_{err} (error bars $\sigma \rightarrow f_{\text{err}}\sigma$). The right-hand column shows these quantities as a function of the photometric error relative to the peak-to-peak amplitude, $\sigma/2A$. A point marks the position of the χ^2/dof for OGLE-LMC-RRLYR-00001, and a square marks its model preference.

is favoured even though the light curve is periodic. If we increase f_{err} to $f_{\text{err}} \approx 1.4$, we can make $\chi^2/\text{dof} \simeq 1$, $\tau \rightarrow 10^4$ d, and have a preference for the periodic model. The RR Lyra behaves similarly to the sawtooth, but would need $f_{\text{err}} \simeq 2$ to achieve $\chi^2/\text{dof} = 1$. The non-sinusoidal light-curve shape requires the assumption of additional white noise (higher values of f_{err}) for the preferred solution to be periodic.

In the second experiment, we generated light curves with different ratios $\sigma/2A$ between the errors and the peak-to-peak variability amplitude, $2A$. Figs 5(c) and (d) show the best-fitting coherence ratio, τ/P , the goodness of fit χ^2/dof and the degree of preference for the QPO or periodic models for this case. The sinusoidal model behaves exactly as expected, always finding the limit $\tau \rightarrow 10^4$ d, with $\chi^2/\text{dof} \simeq 1$, and an overwhelming preference for the periodic solution. The sawtooth model, on the other hand, prefers low-coherence QPO models at high signal-to-noise ratio and then switches to preferring the periodic solution at low signal-to-noise ratio. In the transition between these two limits at intermediate signal-to-noise ratio, the fits to the data are poor, with χ^2/dof

significantly greater than unity. OGLE-LMC-RRLYR-00001, whose parameters are indicated with an ‘x’, behaves like a sawtooth light curve.

We can explain these behaviours by thinking about the problem in terms of the power spectrum. The QPO model produces a Lorentzian power spectrum $P(\omega) = \sigma^2 / \sqrt{8\pi} \sum_{\pm} [\tau^{-2} + (\omega \pm \omega_0)^2]^{-1}$ for a period of $P = 2\pi/\omega_0$ and a damping time-scale τ . A sinusoidal light curve becomes a pair of delta functions at $\pm\omega_0$. A white noise spectrum is simply a constant independent of ω . The spectrum of a non-sinusoidal periodic variable is a series of delta functions at frequencies $\pm n\omega_0$, $n = 1, 2, \dots$, with the envelope of amplitudes representing the Fourier series representation of the light-curve shape.

When we fit a periodic triangle wave in the periodic limit of the QPO, we can match only the first, $n = 1$ peak at $|\omega| = \omega_0$ in the power spectrum, leaving all the power in the $n > 1$ peaks. We can obtain a better fit by increasing the error estimates, which corresponds to adding additional white noise. This works best when the signal-to-noise ratio of the triangle wave is low because white noise is not a very good representation of the $n > 1$ peaks. For the

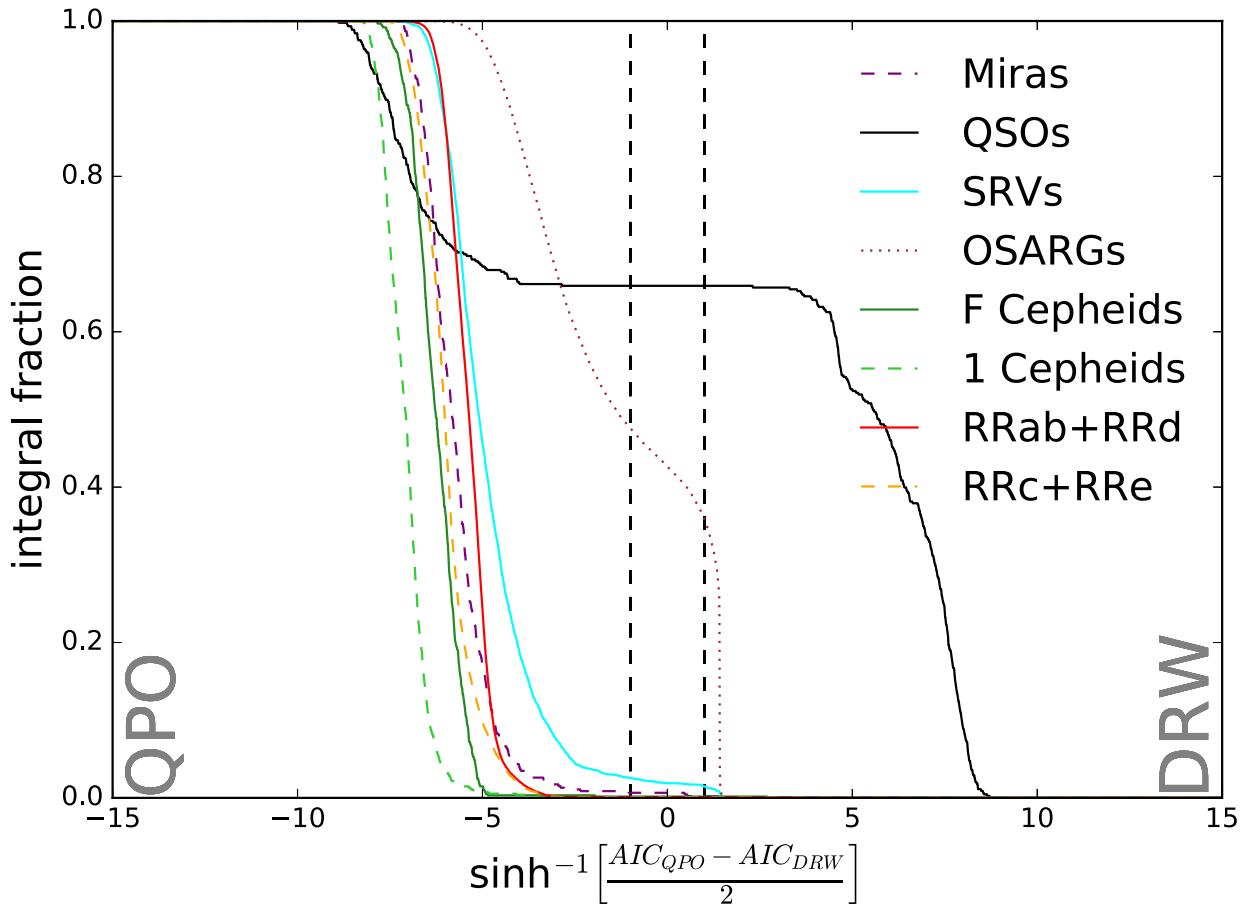


Figure 6. The integral distribution of the labelled variable classes in $\sinh^{-1}[(\text{AIC}_{\text{QPO}} - \text{AIC}_{\text{DRW}})/2]$, where the QPO (DRW) model is favoured for negative (positive) values. The $\sinh^{-1}x$ mapping is linear for $x \ll 1$ and then logarithmic for large x , so that the full dynamic range of the likelihood ratio can be displayed. The mapping is normalized so that the two vertical lines at $|\sinh^{-1}x| = 1$ correspond to 95 per cent confidence likelihood ratios in favour of the QPO ($\sinh^{-1}x = -1$) or DRW ($\sinh^{-1}x = 1$) models. The BIC distributions are very similar.

noise fixed to the correct level, the power in the extra peaks can also be partially captured by giving the model a finite τ . Suppose, for example, that the ratio of the power in the $n = 1$ and $n = 2$ peaks is R_2 , then the QPO model can have the same power at the second peak if $\tau \sim P R_2^{-1/2}$. So the more the light-curve shape deviates from truly sinusoidal, the more the best-fitting model will prefer a finite τ even though the underlying model is actually periodic.

This strongly suggests testing a higher order QPO model with the second period fixed to be $P/2$ (i.e. the $n = 2$ term). Periodic variables, such as Cepheids and RR Lyrae, with strongly non-sinusoidal light curves, can be well fit by Fourier series with period-dependent amplitudes (see e.g. Pejcha & Kochanek 2012) and therefore should be better fit by a QPO model with an $n = 2$ term. A simple second-order example of such a light curve, $f(t) = A \cos(2\pi t/P) + B \cos(4\pi t/P + \phi)$, corresponds to the periodic limit of a higher order ($p = 4$) stochastic process with the periods fixed in a 2:1 ratio. In Section 4, we will compare QPO results for the single component ($p = 2$) model to the higher order $p = 4$ QPO model results for RR Lyrae and Cepheids with the primary period fixed to the primary OGLE period and the secondary Fourier period fixed to half this period, fitting only an amplitude of oscillation for the primary and secondary Fourier periods, and a shared coherence time (A, B and τ). The phase ϕ does not affect the autocorrelation function.

4 RESULTS

Keeping these issues in mind, we now examine the results from fitting all the variables with the DRW and QPO models. We first explore whether the QPO model is a better description of periodic and quasi-periodic phenomena than the DRW model. Next, we examine how the QPO model distinguishes between periodic and quasi-periodic phenomena. Finally, we examine the QPO parameter distributions for the various variable classes with an eye towards classification.

4.1 Discriminating between the QPO and DRW models

Fig. 6 shows the distribution of the variable classes in the relative probabilities of the QPO and DRW models for the AIC likelihoods. We use the $\sinh^{-1}x$ scaling of equation (6) to compress the full dynamic range into the figure – when $|\sinh^{-1}x| = 1$ the associated model is favoured at 95 per cent confidence.

Quasar variability is generally believed to be stochastic rather than quasi-periodic, and we find that a majority of the quasars (58 per cent) strongly favour the DRW model ($f > 1$), leaving 42 per cent of the QSOs that prefer the QPO model. This is broadly consistent with the earlier study of Andrae et al. (2013), who found that the DRW best described 80 per cent of the 6304 QSOs in their

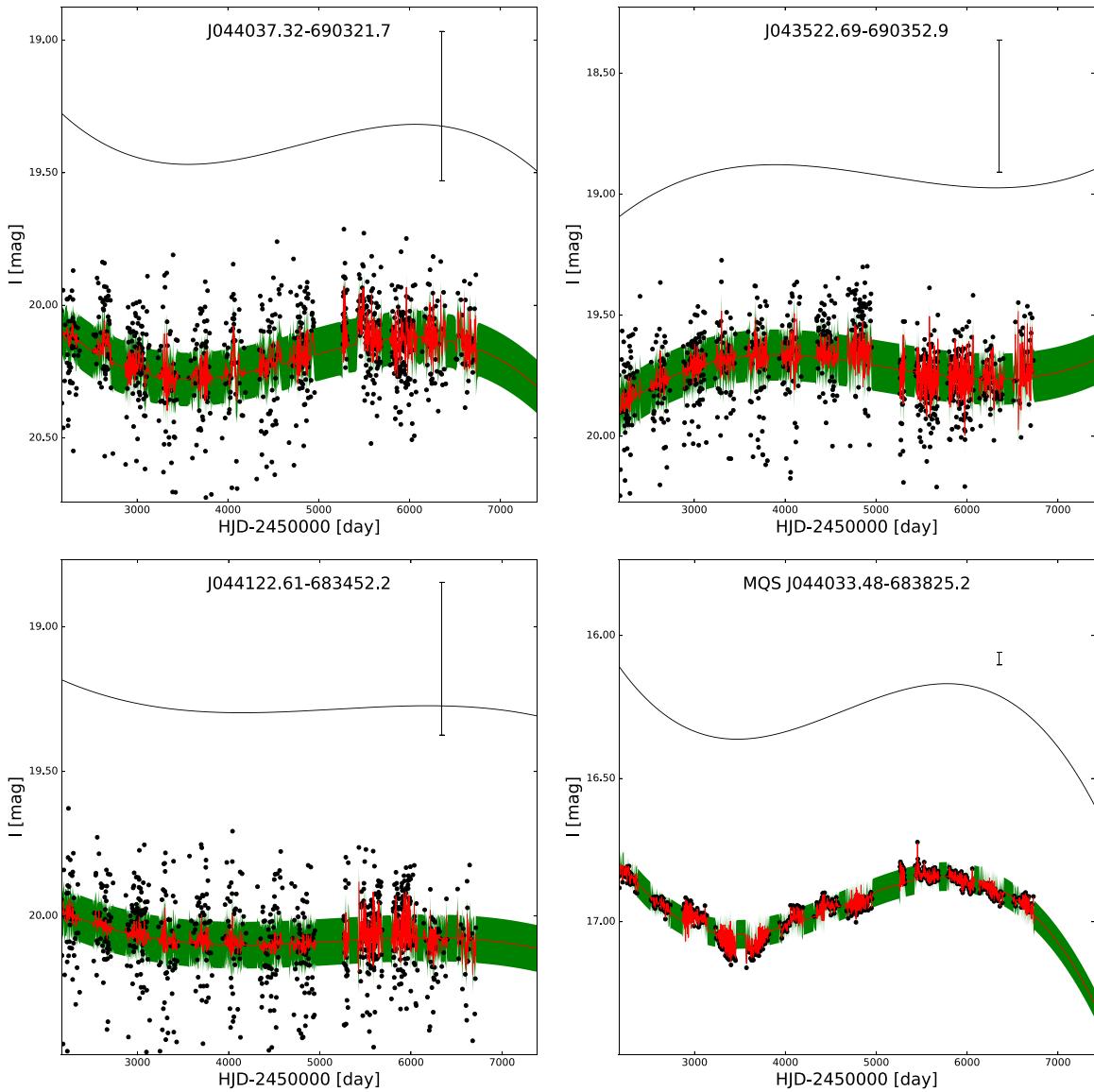


Figure 7. Four examples of QSOs for which the QPO model was favoured over the DRW model. The points are the OGLE data and the error bar indicates the median photometric error. The offset curves show the cubic trend models. The curve and shaded regions again correspond to the average of the best-fitting light curves and their dispersion (see Section 3). All these examples have QPO periods smaller than the OGLE observing cadence, and are nominally coherent, but with a coherence time that is also smaller than the OGLE cadence. Essentially, the cubic trend model has captured the true variability and the QPO model is acting as a contribution to the noise.

sample, while 27 per cent preferred a simultaneous DRW and sinusoid model. While the authors did not consider the CARMA(2,1) QPO model, their finding that a significant fraction was consistent with a purely periodic sinusoid plus a DRW seems consistent with our results. Fig. 7 shows four examples of QSO light curves that strongly favour the QPO model. They are all quasars where the light curves are dominated by long-term trends over the time span of the data, with 22 per cent also having very poor signal-to-noise ratios. One problem with the MQS quasar sample is that many of the quasars are relatively faint for the OGLE survey, and so lack high-quality light curves.

In reality, the preference of most quasars for the QPO model is illusory because the best-fitting QPO model is also ‘incoherent’ in the sense that $\tau/P < 1$. Essentially a cubic trend model plus a long period, relatively incoherent oscillation can fit slow quasar

variability well in such cases. If we add the requirement that the QPO model must not only be preferred by the information criterion, but must also have a best-fitting QPO model with $\tau/P > 1$, then the fraction of ‘quasi-periodic’ quasars drops below 20 per cent. Even among these objects, however, half have a coherence time less than the OGLE cadence. In this regime, the QPO model differs little from the white noise limit of a DRW model: the period is meaningless because the coherence time is so short that its effect is to broaden the photometric errors. These solutions are found in spite of evident variability on time-scales greater than the cadence. Essentially the variability is largely modelled by the cubic trend, and the QPO is acting as an additional source of white noise (see e.g. Fig. 7).

As one would expect, most of the variable stars in Fig. 6 are better fit by the QPO model. RR Lyrae and Cepheids overwhelmingly show a strong preference ($f < -1$) for the QPO over the DRW model.

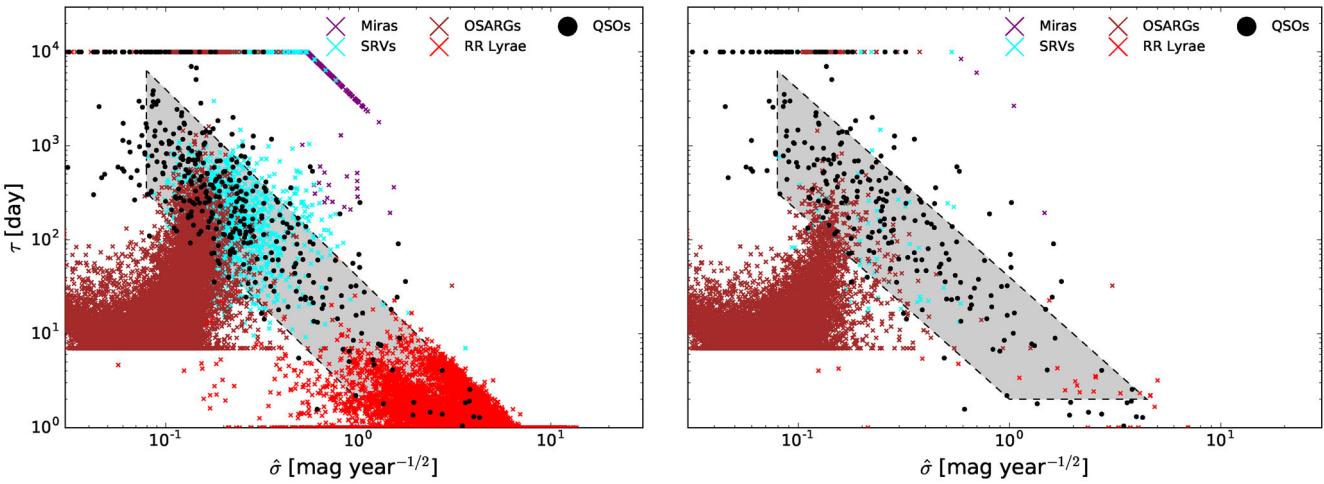


Figure 8. Distribution of several variable classes and QSOs in the DRW parameter space of $\hat{\sigma} \equiv \sigma_{\text{DRW}} \sqrt{2/\tau}$ and τ to match the equivalent figures in Kozłowski et al. (2010). The left-hand panel shows all objects, while the right-hand panel shows the distribution after eliminating sources that prefer ($f < 0$ in Fig. 6) the QPO model. The grey shaded region is the quasar selection region proposed by Kozłowski et al. (2010). Given the uncertainties in estimating τ , the optimal selection region for QSOs will vary based on the time baseline of a typical light curve. For our purposes, we retain the selection region based on OGLE-III time baselines. Imposing the cut on the relative likelihoods of the DRW and QPO models eliminates ~ 7 per cent of the quasars but 67 per cent of the variables.

The LPVs prefer a QPO over the DRW in similar proportions. A small fraction of the RR Lyrae (0.02 per cent) and Cepheids (0.1 per cent) have a strong preference ($f > 1$) for DRW. Only 0.4 per cent of RR Lyrae and 0.3 per cent of Cepheids are incoherent ($\tau/P < 1$). The results with a higher order QPO model introduced in Section 3 are similar, but the higher order models eliminate the ‘tail’ of systems towards a preference for the DRW model, particularly for RRab and RRd sources whose light curves are not well described by sinusoids.

The largest population of DRW-preferring variable stars is the OSARGs. Almost a quarter of the OSARGs in our sample prefer the DRW model. The best-fitting QPO models for the OSARGs better modelled by the DRW model almost always have $\tau/P < 1$. In the limit of incoherent oscillations, the QPO model increasingly resembles the DRW model, and the fits become indistinguishable by their likelihoods and χ^2/dof . In this limit, the AIC model preference will always favour a DRW model, because it has fewer parameters. Although periods may remain detectable in a periodogram for marginally incoherent sources, the DRW model can still be statistically preferred in the AIC formalism due to its fewer degrees of freedom. An issue to keep in mind for the OSARGs is that their variability amplitudes are comparable to their photometric noise, which makes it difficult to distinguish among models.

Fig. 8 reprises the similar figure in Kozłowski et al. (2010), comparing the distribution of quasars and variable stars in DRW parameters. If no other criteria are imposed (left-hand panel), the selection region proposed by Kozłowski et al. (2010) contains 62 per cent of the quasars, but also contains 0.52 variable stars for every quasar.

In practice, Kozłowski et al. (2010) considered additional simple selection criteria (e.g. colour and magnitude) that significantly increase the efficiency of the selection process, but here we only consider the time domain criteria. If we impose the criterion that the light curves must also prefer the DRW model, so $\text{AIC}_{\text{QPO}} - \text{AIC}_{\text{DRW}} > 0$, the contamination is greatly reduced (the right-hand panel of Fig. 8). The selection region now contains 55 per cent of the quasars, but only 0.20 variable stars for every quasar. The additional criterion loses only 7 per cent of the quasars but reduces the stellar contamination by 67 per cent. We

attempted to increase the efficiency of quasar selection by also selecting objects that had incoherent QPO models. However, this does not improve quasar selection because the largest contaminant is OSARGs, which also favour incoherent QPO models. The MQS quasars are, of course, selected from a region of very high stellar density (the Magellanic Clouds!), which makes the density of variable stars tremendously larger than in a typical extragalactic field (~ 1 QSO per 10^5 stars in the Magellanic Clouds!). In a true extragalactic field, the purity of such a variability-selected sample would be far higher. Additional criteria such as the magnitude and colour criteria considered by Kozłowski et al. (2010) could also be developed to avoid the issues leading to quasars favouring the QPO model. For example, we could also require more variability ‘power’ in the stochastic process than in the cubic polynomial or require a minimum variability power.

4.2 Periodic versus quasi-periodic

We next test whether sources are better fit as truly periodic (i.e. $\tau \rightarrow \infty d$) or as quasi-periodic (i.e. finite τ) variables by comparing the likelihoods for a QPO with fixed $\tau/P = 10^4$ (as a proxy for $\tau \rightarrow \infty d$) to those with τ free to vary, as shown in Fig. 9. Quasars, as we would expect, strongly favour the QPO model, with 97 per cent preferring ($f < 0$) the QPO model. Among the few exceptions, two have extreme outliers in the light curves (up to 25σ) and one has an exceptionally short light curve of 74 epochs. As discussed earlier, when quasars are modelled as QPOs they strongly favour either the ‘incoherent’ limit ($\tau/P < 1$) where the QPO model becomes increasingly similar to the DRW model or the white noise limit where τ is smaller than the cadence.

It is not surprising that the LPVs shown in Fig. 9 generally favour a QPO over a periodic model, as these variables generally do not have simple, coherent light-curve structures. Only 25 per cent of the Miras strongly prefer ($f > 1$) a periodic model, and most of the Miras that prefer a periodic model have genuinely regular, sinusoidal light curves with mildly variable (~ 10 per cent) amplitudes. The less regular SRVs show a smaller periodic fraction while the low-amplitude OSARGs show an overwhelming preference for the

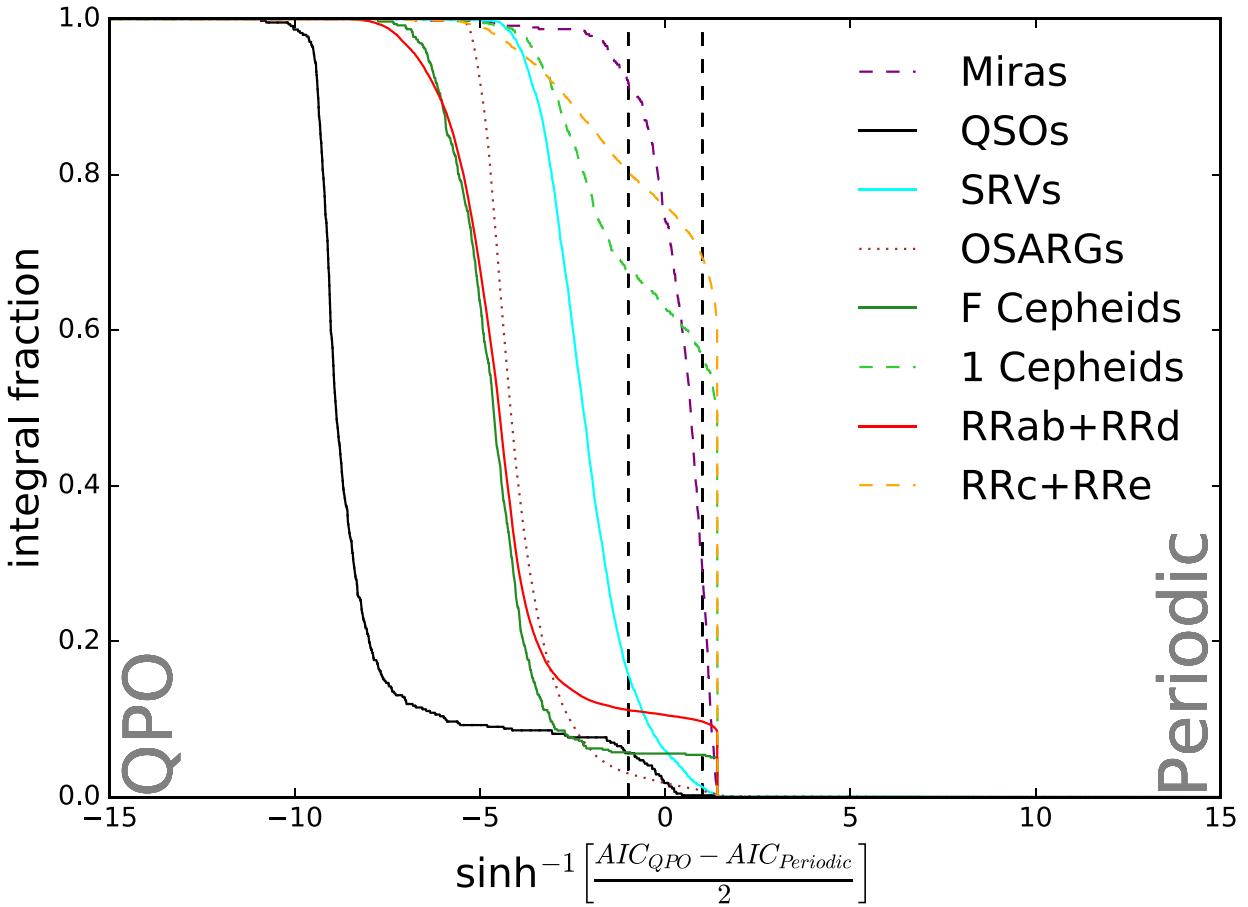


Figure 9. Distribution of selected variables in the AIC likelihood ratio $\sinh^{-1} x$ with $x = (\text{AIC}_{\text{QPO}} - \text{AIC}_{\text{P}})/2$ for favouring periodic ($\sinh^{-1} x > 0$, right) or quasi-periodic ($\sinh^{-1} x < 0$, left) variability.

QPO model. The low amplitudes of the OSARGs also lead to noisier light curves, where it becomes difficult to distinguish the models (see Fig. 5 and the discussion in Section 3). The LPVs also have relatively low coherence ratios, with ~ 20 per cent of Miras and less than 1 per cent of SRVs and OSARGs having $\tau/P > 10$.

The surprise, leading to much of the discussion in Section 3, is that a large fraction of periodic RR Lyrae and Cepheid variables apparently favour a finite coherence time. For Cepheids, more than 30 and 90 per cent of first overtone and fundamental Cepheids, respectively, strongly prefer the QPO model even though they are very periodic oscillators. RR Lyrae shows a similar trend, with 20 and 85 per cent of RRc and RRab, respectively, preferring a QPO. This is not due to the presence of Blazhko effect RR Lyrae alone, although such Blazhko RR Lyrae do tend to prefer the QPO model (see below). For those objects that do prefer the QPO model, 95 per cent of RRc have coherence ratios $\tau/P > 100$, whereas 95 per cent of RRab have $\tau/P < 100$. Cepheids show a similar pattern, with more than 80 per cent of fundamental mode Cepheids favouring a coherence ratio $\tau/P < 10$, while more than 95 per cent of first overtone Cepheids have $\tau/P > 10$. As discussed in Section 3, these behaviours are a consequence of light-curve shape. First overtone Cepheids, RRc and RRe variables all have relatively sinusoidal light curves, while fundamental mode Cepheids, RRab and RRe variables have more triangular/sawtooth waveforms. Qualitatively, the integral fraction curves in Fig. 9 favour periodic models more

strongly when the light curve is sinusoidal and favour QPO models more strongly when the light curves are sawtooth like.

We introduced the higher order $p = 4$ QPO models with periods locked in a 2:1 ratio and a common coherence time to explore if this difference was driven by the assumption of sinusoidal waveforms in the one-component QPO model. We show in Fig. 10 that using the two-component QPO models halves the number of RRab and RRe variables that prefer a QPO compared to the one-component model. The increase in periodic preference going from a one-component to the two-component QPO models for the more sinusoidal RRc and RRe sources is marginal. An even larger shift is seen for the fundamental mode Cepheids. These results are in agreement with the behaviour we found in Section 3 for the experiments comparing artificial light curves with sawtooth or sinusoid waveforms. It is not surprising, therefore, that a higher order QPO model results in a greater preference for periodicity. The two-component QPO is a better approximation to the sawtooth-like light curves of RRab, RRe and fundamental mode Cepheids, so the coherence time can act more like the desired physical parameter, and less as a proxy for light-curve shape.

In addition to light-curve shape, we also noted in Section 2 that mischaracterized errors can bias the results, irrespective of light-curve shape. If some of the noise level assumed in the model is less than the true noise, the extra variance will be interpreted as signal. The natural way to model this extra variance is to move

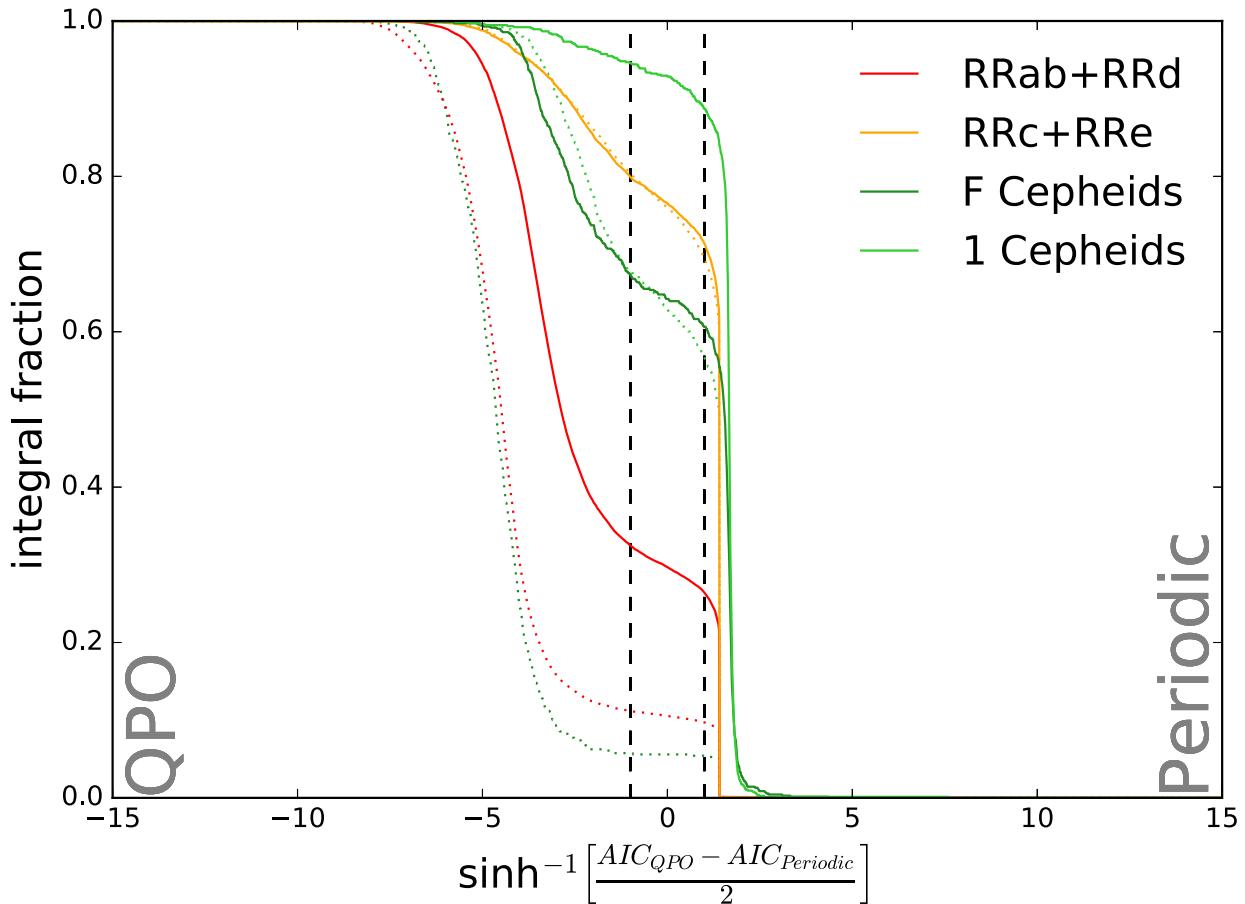


Figure 10. Distribution of the RRab+RRd, RRc+RRe, fundamental Cepheids and first overtone Cepheids in the AIC likelihood difference for the QPO and periodic models when using either the standard QPO model (dotted) or the higher order model (solid). The changes are small for the more sinusoidal RRc+RRe and first overtone Cepheids, but shift strongly towards the periodic case for the more sawtooth-like RRab+RRd and fundamental mode Cepheids.

the model away from periodic to model the mischaracterized noise as stochastic variability. If we take the best-fitting model and re-estimate the errors so that it has $\chi^2/\text{dof} = 1$, then the fraction of Cepheids and RR Lyrae favouring the periodic model increase significantly, from 70 and 10 per cent to 90 and 65 per cent in the case of first overtone and fundamental Cepheids, respectively, and increasing from 80 and 15 per cent among RRc/RRe and RRab/RRd to 95 and 90 per cent, respectively.

4.3 Variable distributions in QPO parameters

The lesson from Section 3 is that the QPO τ parameter cannot simply be interpreted as a measure of coherence because it is significantly affected by light-curve shape. This is, however, no barrier to using the QPO parameters to classify variables. Fig. 11 shows the distribution of all the sources in the $P-\tau$, $P-\sigma$ and $\tau-\sigma$ QPO parameter spaces. We see that the $P-\tau$ plane maps physically distinct variable types to nearly unique spaces. Below we briefly discuss the properties of each class of variable. We do not examine the properties of the variables in DRW parameter space since such an exploration was carried out in Kozłowski et al. (2010).

If we first consider the distribution of the quasars, we see that they largely have long periods ($P \gtrsim 100$ d) and low coherence ($\tau/P < 1$). There are also a small number with short periods, small coherence

time and nominally higher coherence, as discussed in Section 4.1 and Fig. 7. QSOs overlap with part of the locus of the OSARGs (discussed below), and therefore cuts in QPO space alone are not sufficient to separate quasars from these stellar variables. However, cuts in QPO/DRW model preference, as proposed in Section 4.1, are efficient.

The various LPVs generally occupy somewhat different regions of the QPO parameter space. That they have differences in period and amplitude is, of course, already known, so we focus on the new parameter τ . Fig. 12 shows the distribution of the Miras and SRVs in $P-\tau$ space further divided by their C/O abundance class. While not perfectly separated by τ at fixed period, it is generally true that the Miras are more coherent (larger τ/P) than the SRVs and that C-rich SRVs are more coherent than O-rich SRVs. Similarly, at a fixed amplitude σ , the coherence time separates the O- and C-rich SRVs well and provides some discrimination between the O- and C-rich Miras. It is probably not possible to well separate the OSARGs from the other LPVs simply based on their QPO parameters, possibly because there are all simply part of an evolutionary continuum as suggested by Soszyński et al. (2013). While our discussion in Section 3 and the results of Section 4 suggest that interpreting τ simply as a coherence time is risky, it is broadly true that longer period Miras and SRVs have larger τ/P suggesting that they are more coherent oscillators. This is also seen simply from folded

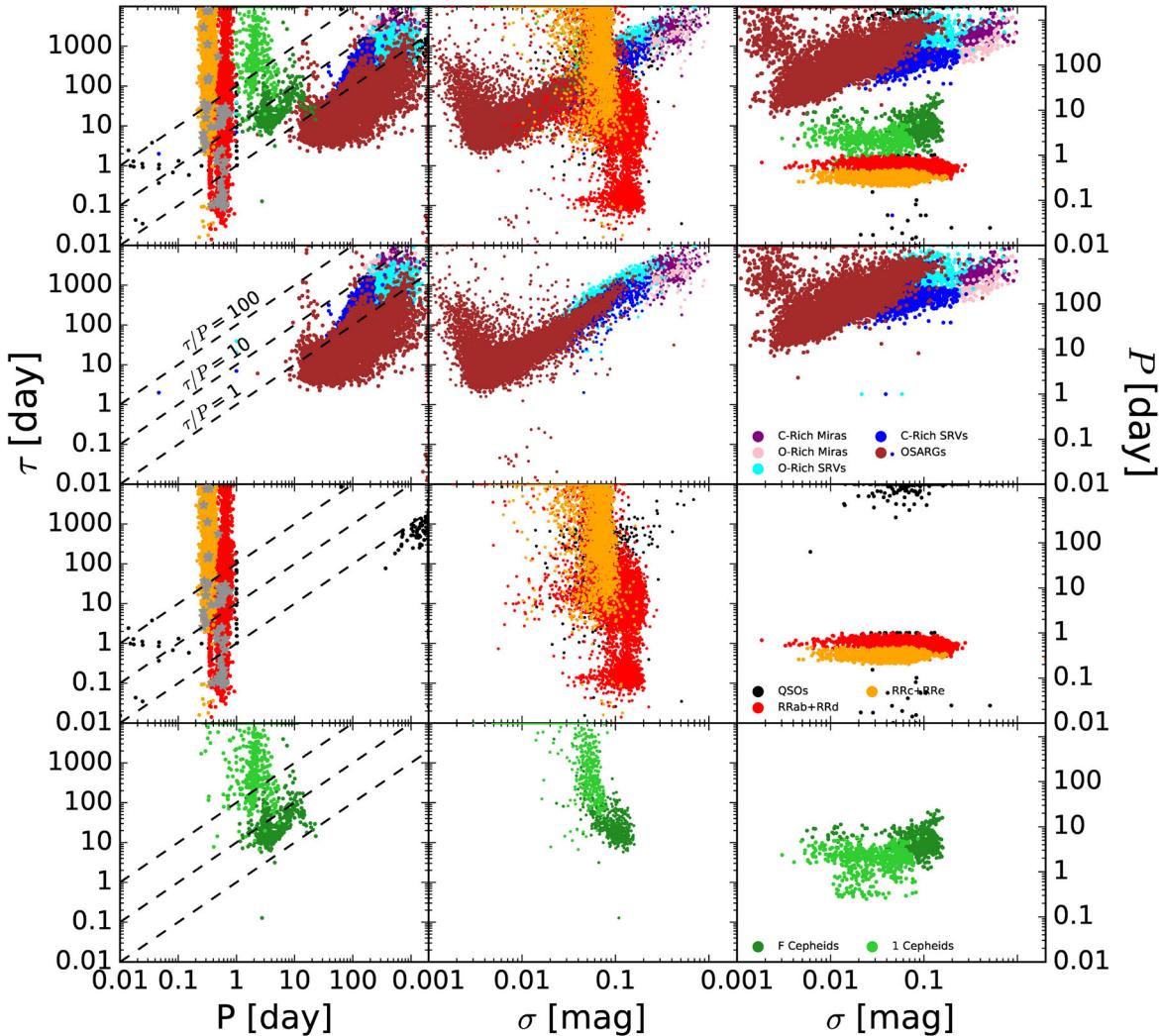


Figure 11. Variable distributions in projections of the $P-\sigma-\tau$ QPO parameters. Note that chemical distinctions are included for SRVs and Miras. In the $P-\tau$ plane, dotted lines of constant coherence, τ/P , are shown to guide the eye, where $\tau/P = 1$ is the boundary between coherent and incoherent periodicity. The first two columns show τ on the y-axis, whereas the third column is against P . Grey stars indicate RR Lyrae classified as Blazhko by Chen, Jiang & Yang (2013).

light curves, where longer period Miras look more like the upper panels of Fig. 1 and shorter period ones look more like the lower panels.

Although they are highly regular variables, the Cepheids are not found to have large coherence times τ in the QPO models for all the reasons discussed earlier. That τ appears primarily to be a measure of light-curve shape rather than coherence of oscillation does lead to a clear separation of the fundamental mode and overtone Cepheids at fixed period. Similarly, the RRc and RRe variables tend to have larger τ than the RRab and RRe variables, but these classes are so well separated in period that the QPO parameter adds little new information. Blazhko effect RR Lyrae should appear as low-coherence QPOs, although they are too rare to explain the overall distributions. If we look at the parameters of the 52 confirmed Blazhko RR Lyrae from Chen et al. (2013), we see that they are located preferentially at lower coherence ratios among RRab but are nearly uniform in distribution for RRc and RRe. With a higher order model that would make all the normal RR Lyrae coherent oscillators, the coherence time would likely be an efficient means of identifying Blazhko variables. In particular, Fig. 13 shows the

$P-\tau$ plane for the RR Lyrae and Cepheid variables using the $p = 4$ QPO model. Both RR Lyrae and Cepheids now have significantly higher coherence times, as expected from the discussion in Section 3, and the Blazhko RR Lyrae are better separated.

5 DISCUSSION

We have carried out a large-scale survey of the properties of quasars and variable stars using two different models of stochastic variability. The models are the two lowest order models within the CARMA family of solutions to stochastic differential equations: (1) the DRW model characterized by an amplitude σ and a damping time τ and (2) the QPO model characterized by an amplitude σ , a period P and a coherence time τ . In general, variable stars are better described by the QPO model while quasars are better described by the DRW model. This is not a claim that the DRW is a perfect model of QSO variability, although it is completely adequate for our present purposes. There are some ambiguities in the details of the model classification because the QPO model becomes indistinguishable from the DRW model in the limits where the oscillations are very

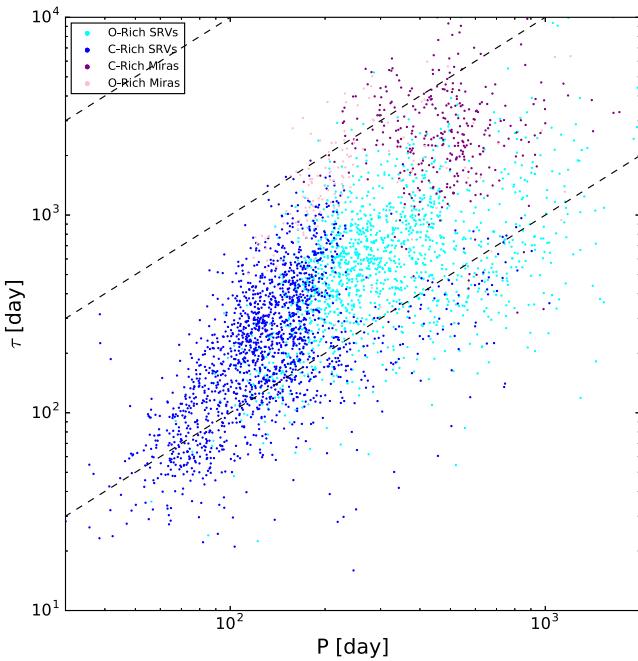


Figure 12. A subset of the top left-hand panel of Fig. 11 showing only Miras and SRVs.

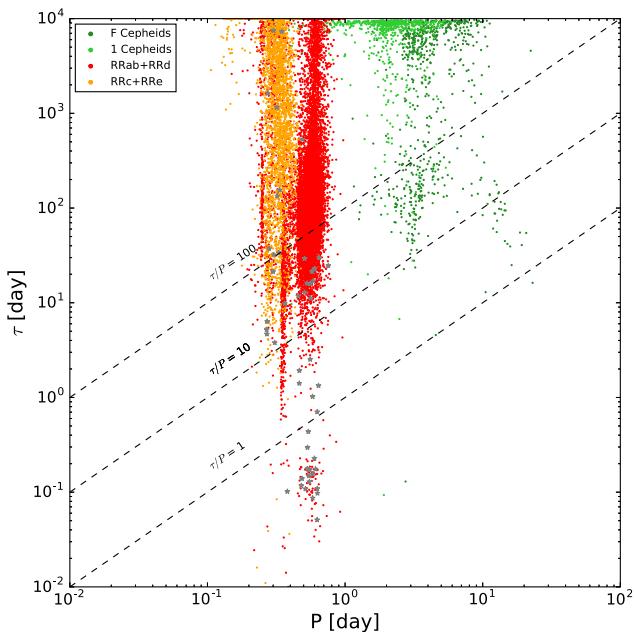


Figure 13. The $P-\tau$ plane for RR Lyrae and Cepheids using the higher order $p = 4$ QPO model. Grey stars indicate RR Lyrae classified as Blazhko by Chen et al. (2013).

incoherent ($\tau/P < 1$), the period approaches the time span of the data or the coherence time, τ , becomes shorter than the observing cadence. By combining the results from the two stochastic models, we can significantly increase the efficiency in separating quasars and variable stars over using the DRW model results alone.

In principle, a quantity like the coherence time in the QPO models has the potential to provide important physical information on the nature of pulsations or, if applied to spotted stars, to the evolution time-scale of the spots. We found, however, that the estimated value

of τ in the QPO model is also quite sensitive to the deviations of the light-curve shape from a sinusoid. Qualitatively, this is most easily understood in terms of the power spectrum of the variability. The QPO model has a power spectrum that is a Lorentzian centred on the principle frequency with a width set by the damping time-scale. For a variable with significant deviations from a sinusoid, such as the ‘sawtooth-like’ long-period Cepheids or RRab variables, the best $p = 2$ QPO models have a finite coherence time in order to broaden the Lorentzian to capture some of the variability power in the overtones of the fundamental period. If we use a higher order QPO model with terms at both the fundamental period and the first overtone, in order to better mimic the shape of such light curves, we find a longer coherence time. Thus, for the coherence time-scale of these models to have that physical meaning, the model must be of high enough order to be able to match the shape of the underlying light curve. This is related to the models of a few individual variable sources by Kelly et al. (2014), where they continued increasing the order of the stochastic model until no increase in the order was statistically justified. For the RRab they considered, they found that a higher order CARMA(7, 0) model fit the light curve best. Kelly et al. (2014) did not force any fixed period ratios between the terms, but since many variables are well described by Fourier series (e.g. Simon & Lee 1981; Simon & Teays 1982), the approach we tested of adding terms with fixed period ratios should work extremely well for many variables.

While the QPO coherence time does not have a simple physical interpretation because of these concerns, the QPO parameter space still provides a means of separating variable classes by adding an additional parameter beyond a period and an amplitude. As noted above, quasars tend to occupy regions of QPO parameter space distinct from variable stars. At fixed period, fundamental and first overtone Cepheids have different coherence times, as do O- and C-rich Miras and SRVs. While not a panacea for classification, models of variable sources combining the results of DRW and QPO models provide a relatively straightforward approach to providing quantitative measures of light curves that can be used for classification. Furthermore, the model fits can all be carried out in $O(N)$ operations using the algorithm discovered by Ambikasaran et al. (2014) even when evaluating the full model likelihood rather than simply using the forecasting approach of Kelly et al. (2014).

Finally, the combination of the DRW and QPO models provides a systematic approach to evaluating periodicity in quasars. There is considerable interest in identifying periodic behaviour in quasar light curves as a probe of quasar binaries and there are now a number of claimed detections (e.g. Sillanpaa et al. 1988; Maness et al. 2004; Rodriguez et al. 2006; Eracleous et al. 2012; Graham et al. 2015; Liu et al. 2015; Yan et al. 2015; Bhatta et al. 2016; Li et al. 2016; Zheng et al. 2016). The problem is that standard means of evaluating periodicity, such as a Lomb–Scargle periodogram (Lomb 1976; Scargle 1982), are essentially comparing the probability of modelling the source as a sinusoid to the probability of the same light curve being generated by a known level of white noise characterized by the photometric errors. As discussed by Press (1978) and more recently by Vaughan et al. (2016), this is not a good statistical test when the underlying source class is known to be variable and with a relatively red noise-like power spectrum showing more variability on longer time-scales (for a quasar binary search taking in to account red noise, however, see Charisi et al. 2016). We see this in our analyses of the MQS quasars, where effects like noise levels or the structure of the variability can make the QPO model of a quasar more likely than the DRW model. The QPO model we consider here, which spans variability from truly periodic back to the

completely non-periodic DRW model, provides a good statistical basis for evaluating periodicity in quasar light curves (modulo the caveats above) because it has a continuous parameter τ to evaluate the significance of the periodicity. Our approach also has the ability to systematically remove light-curve means or other calibration issues for light curves combining multiple sources of data.

ACKNOWLEDGEMENTS

CSK is supported by NSF grant AST-1515927. We would like to thank Professor M. Kubiak and G. Pietrzyński, former members of the OGLE team, for their contribution to the collection of the OGLE photometric data over the past years. The OGLE project has received funding from the Polish National Science Centre grant MAESTRO no. 2014/14/A/ST9/00121 to AU. SK acknowledges the financial support from the Polish National Science Center grant no. 2014/15/B/ST9/00093.

REFERENCES

- Aigrain S., Parviainen H., Pope B. J. S., 2016, MNRAS, 459, 2408
- Ambikasaran S., Foreman-Mackey D., Greengard L., Hogg D. W., O’Neil M., 2014, preprint ([arXiv:1403.6015](https://arxiv.org/abs/1403.6015))
- Andrae R., Kim D.-W., Bailer-Jones C. A. L., 2013, A&A, 554, A137
- Auvergne M. et al., 2009, A&A, 506, 411
- Bakos G. Á., Lázár J., Papp I., Sári P., Green E. M., 2002, PASP, 114, 974
- Balaji B., Croll B., Levine A. M., Rappaport S., 2015, MNRAS, 448, 429
- Ball N. M., Brunner R. J., 2010, Int. J. Modern Phys. D, 19, 1049
- Bedding T. R., Zijlstra A. A., 1998, ApJ, 506, L47
- Bedding T. R., Kiss L. L., Kjeldsen H., Brewer B. J., Dind Z. E., Kawaler S. D., Zijlstra A. A., 2005, MNRAS, 361, 1375
- Bhatta G. et al., 2016, ApJ, 832, 47
- Bloom J. S. et al., 2012, PASP, 124, 1175
- Borucki W. J. et al., 2010, Science, 327, 977
- Brockwell P. J., Davis R. A., 1996, Introduction to Time Series and Forecasting, Springer-Verlag, New York
- Brockwell P. J., Davis R. A., 2002, Introduction to Time Series and Forecasting, 2nd edn. Springer-Verlag, New York (<http://www.worldcat.org/isbn/0387953515>)
- Cano Z., Wang S.-Q., Dai Z.-G., Wu X.-F., 2016, preprint ([arXiv:1604.03549](https://arxiv.org/abs/1604.03549))
- Charisi M., Bartos I., Haiman Z., Price-Whelan A. M., Graham M. J., Bellm E. C., Laher R. R., Márcia S., 2016, MNRAS, 463, 2145
- Chen B.-Q., Jiang B.-W., Yang M., 2013, Res. Astron. Astrophys., 13, 290
- Christensen-Dalsgaard J., Kjeldsen H., Mattei J. A., 2001, ApJ, 562, L141
- Christy R. F., 1966, ApJ, 144, 108
- Debosscher J., Sarro L. M., Aerts C., Cuypers J., Vandenbussche B., Garrido R., Solano E., 2007, A&A, 475, 1159
- Dexter J., Agol E., 2011, ApJ, 727, L24
- Drake A. J. et al., 2009, ApJ, 696, 870
- Eracleous M., Boroson T. A., Halpern J. P., Liu J., 2012, ApJS, 201, 23
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
- Frasca A., Fröhlich H.-E., Bonanno A., Catanzaro G., Biazzo K., Molenda-Zakowicz J., 2011, A&A, 532, A81
- García R. A. et al., 2014, A&A, 572, A34
- Graczyk D. et al., 2011, Acta Astron., 61, 103
- Graham M. J., Djorgovski S. G., Drake A. J., Mahabal A. A., Chang M., Stern D., Donalek C., Glikman E., 2014, MNRAS, 439, 703
- Graham M. J. et al., 2015, MNRAS, 453, 1562
- Hamuy M., Phillips M. M., Suntzeff N. B., Schommer R. A., Maza J., Smith R. C., Lira P., Aviles R., 1996, AJ, 112, 2438
- He S., Yuan W., Huang J. Z., Long J., Macri L. M., 2016, AJ, 152, 164
- Howarth J. J., 1991, J. British Astron. Assoc., 101, 101
- Ivezic Z. et al., 2008, preprint ([arXiv:0805.2366](https://arxiv.org/abs/0805.2366))
- Kaiser N. et al., 2002, in Tyson J. A., Wolff S., eds, Proc. SPIE Vol. 4836, Survey and Other Telescope Technologies and Discoveries. SPIE, Bellingham, p. 154
- Kasliwal V. P., Vogeley M. S., Richards G. T., 2015, MNRAS, 451, 4328
- Kato T., Uemura M., Ishioka R., Nogami D., Kunjaya C., Baba H., Yamaoka H., 2004, PASJ, 56, S1
- Kelly B. C., Bechtold J., Siemiginowska A., 2009, ApJ, 698, 895
- Kelly B. C., Becker A. C., Sobolewska M., Siemiginowska A., Uttley P., 2014, ApJ, 788, 33
- Kholopov P. N., Samus N. N., Kazarovets E. V., Perova N. B., 1985, Inf. Bull. Var. Stars, 2681, 1
- Kim D.-W., Protopapas P., Bailer-Jones C. A. L., Byun Y.-I., Chang S.-W., Marquette J.-B., Shin M.-S., 2014, A&A, 566, A43
- Klebesadel R. W., Strong I. B., Olson R. A., 1973, ApJ, 182, L85
- Koen C., 2005, MNRAS, 361, 887
- Koen C., 2012, MNRAS, 419, 1197
- Kozłowski S., 2017, A&A, 597, A128
- Kozłowski S. et al., 2010, ApJ, 708, 927
- Kozłowski S. et al., 2013, ApJ, 775, 92
- Law N. M. et al., 2009, PASP, 121, 1395
- Li Y.-R. et al., 2016, ApJ, 822, 4
- Liu T. et al., 2015, ApJ, 803, L16
- Lomb N. R., 1976, Ap&SS, 39, 447
- LSST Science Collaboration et al., 2009, preprint ([arXiv:0912.0201](https://arxiv.org/abs/0912.0201))
- Mackenzie C., Pichara K., Protopapas P., 2016, ApJ, 820, 138
- MacLeod C. L. et al., 2010, ApJ, 721, 1014
- MacLeod C. L. et al., 2012, ApJ, 753, 106
- McQuillan A., Aigrain S., Mazeh T., 2013, MNRAS, 432, 1203
- Maness H. L., Taylor G. B., Zavala R. T., Peck A. B., Pollack L. K., 2004, ApJ, 602, 123
- Mushotzky R. F., Edelson R., Baumgartner W., Gandhi P., 2011, ApJ, 743, L12
- Pejcha O., Kochanek C. S., 2012, ApJ, 748, 107
- Pepper J., Gould A., Depoy D. L., 2004, in Holt S. S., Deming D., eds, AIP Conf. Proc. Vol. 713, The Search for Other Worlds. Am. Inst. Phys., New York, p. 185
- Pojmański G., 1997, Acta Astron., 47, 467
- Pojmański G., 2002, Acta Astron., 52, 397
- Pollacco D. L. et al., 2006, PASP, 118, 1407
- Press W. H., 1978, Comments Astrophys., 7, 103
- Pskovskii I. P., 1977, SvA, 21, 675
- Reinhold T., Reiners A., Basri G., 2013, A&A, 560, A4
- Richards J. W., Starr D. L., Miller A. A., Bloom J. S., Butler N. R., Brink H., Crellin-Quick A., 2012, ApJS, 203, 32
- Ricker G. R. et al., 2014, Proc. SPIE, 9143, 914320
- Robinson E. L., 1976, ARA&A, 14, 119
- Rodriguez C., Taylor G. B., Zavala R. T., Peck A. B., Pollack L. K., Romani R. W., 2006, ApJ, 646, 49
- Rybicki G. B., Press W. H., 1992, ApJ, 398, 169
- Sako M. et al., 2008, AJ, 135, 348
- Sako M. et al., 2011, ApJ, 738, 162
- Sarro L. M., Debosscher J., López M., Aerts C., 2009, A&A, 494, 739
- Scargle J. D., 1982, ApJ, 263, 835
- Shappee B. et al., 2014, Am. Astron. Soc. Meeting Abstr., 223, 236.03
- Sillanpää A., Haarala S., Valtonen M. J., Sundelius B., Byrd G. G., 1988, ApJ, 325, 628
- Simon N. R., Lee A. S., 1981, ApJ, 248, 291
- Simon N. R., Teays T. J., 1982, ApJ, 261, 586
- Smith R. M. et al., 2014, Proc. SPIE, 9147, 914779
- Snyder J. A., 1998, in D’Odorico S., ed., Proc. SPIE Vol. 3355, Optical Astronomical Instrumentation. SPIE, Bellingham, p. 635
- Sokal A. D., 1996, preprint, ([arXiv:astro-ph/9601001](https://arxiv.org/abs/astro-ph/9601001))
- Soszyński I., Udalski A., Kubiak M., Szymański M., Pietrzyński G., Zebrun K., Szewczyk O., Wyrzykowski L., 2004, Acta Astron., 54, 129
- Soszyński I. et al., 2008, Acta Astron., 58, 163
- Soszyński I. et al., 2009a, Acta Astron., 59, 1
- Soszyński I. et al., 2009b, Acta Astron., 59, 239
- Soszyński I., Wood P. R., Udalski A., 2013, ApJ, 779, 167

- The Dark Energy Survey Collaboration, 2005, preprint ([arXiv:astro-ph/0510346](https://arxiv.org/abs/astro-ph/0510346))
- Tyson J. A., 2002, in Tyson J. A., Wolff S., eds, Proc. SPIE Vol. 4836, Survey and Other Telescope Technologies and Discoveries. SPIE, Bellingham, p. 10
- Udalski A., Szymański M. K., Soszyński I., Poleski R., 2008, *Acta Astron.*, 58, 69
- 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
- Wood P. R., 1974, *ApJ*, 190, 609
- Woosley S. E., Weaver T. A., 1995, *ApJS*, 101, 181
- Woźniak P. R., 2000, *Acta Astron.*, 50, 421
- Woźniak P. R., Williams S. J., Vestrand W. T., Gupta V., 2004, *AJ*, 128, 2965
- Wray J. J., Eyer L., Paczyński B., 2004, *MNRAS*, 349, 1059
- Wyrzykowski Ł. et al., 2009, *MNRAS*, 397, 1228
- Yan C.-S., Lu Y., Dai X., Yu Q., 2015, *ApJ*, 809, 117
- York D. G. et al., 2000, *AJ*, 120, 1579
- Zheng Z.-Y., Butler N. R., Shen Y., Jiang L., Wang J.-X., Chen X., Cuadra J., 2016, *ApJ*, 827, 56
- Zhevakin S. A., 1953, preprint, ([arXiv:astro-ph/0510346](https://arxiv.org/abs/astro-ph/0510346))
- Zu Y., Kochanek C. S., Peterson B. M., 2011, *ApJ*, 735, 80
- Zu Y., Kochanek C. S., Kozłowski S., Udalski A., 2013, *ApJ*, 765, 106
- Zu Y., Kochanek C. S., Kozłowski S., Peterson B. M., 2016, *ApJ*, 819, 122

APPENDIX: TREND CORRECTION

The mean of the best-fitting light curves, \hat{s} , approximating the observed light curve, y , may be calculated using the expression derived in Rybicki & Press (1992),

$$\hat{s} = \mathbf{SC}^{-1}(y - \mathbf{L}q), \quad (\text{A1})$$

with variance from the true light curve given by

$$\langle(s - \hat{s})^2\rangle = \mathbf{S} - \mathbf{S}^T \mathbf{C}_\perp \mathbf{S}, \quad (\text{A2})$$

where

$$\mathbf{C}_\perp \equiv (\mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{L} (\mathbf{L}^T \mathbf{C}^{-1} \mathbf{L})^{-1} \mathbf{L}^T \mathbf{C}^{-1})^{-1}. \quad (\text{A3})$$

As previously discussed in Rybicki & Press (1992) and Kozłowski et al. (2010), the CARMA formalism can be generalized to take into account arbitrary trends. This assumes that the data are modelled as

$$\mathbf{y} = \mathbf{s} + \mathbf{n} + \mathbf{L}q, \quad (\text{A4})$$

where \mathbf{s} is the expected signal, described in this case by a Gaussian process, \mathbf{n} is a Gaussian noise term and $\mathbf{L}q$ describes a trend. \mathbf{L} is a $N \times \ell$ matrix, with ℓ representing the number of temporal trends to include in the fit and N is the number of data points. The column vector q is of length ℓ and contains the polynomial coefficients for the trend. In this notation, we can include a mean and linear trend in the model, $q_1 + (t - t_0)q_2$, by setting $q = (q_1, q_2)$ and $L_{i1} = 1$, $L_{i2} = t_i - t_0$.

As another example, to remove jumps within a light curve, like the calibration shifts between the OGLE-III and OGLE-IV QSO light curves, one would again have $q = (q_1, q_2)$, but define \mathbf{L} to be $\mathbf{L} = (1, 0)$ for the M OGLE-III points of the light curve and $\mathbf{L} = (0, 1)$ for the $N - M$ OGLE-IV points of the light curve.

This paper has been typeset from a TeX/LaTeX file prepared by the author.