Draft version December 14, 2018
Typeset using LaTeX **twocolumn** style in AASTeX62

Improving Damped Random Walk parameters for SDSS Stripe82 Quasars with baseline extension with PanStarrs1 data.

Krzysztof L. Suberlak,[1] Željko Ivezić,[1] and Chelsea MacLeod[2]

[1] *Department of Astronomy*
University of Washington
Seattle, WA 98195, USA
[2] *Harvard Smithsonian Center for Astrophysics*
60 Garden St, Cambridge, MA 02138, USA

(Received January 1, 2019; Revised January 17, 2019; Accepted February 1, 2019)

Submitted to ApJ

ABSTRACT

1. INTRODUCTION

Quasars are variable. Their light curves have been successfully described using the Damped Random Walk model (Kelly+2009, Macleod+2010, Kozlowski+2010,Zu+2011, Kasliwal+2015,) . The origin of variability is debated, with thermal origin being the favorite explanation (Czerny+1999, Kelly+2013,), connected to the inhomogeneity of the accretion disk (Dexter,Agol 2011), or even magnetically elevated disks (Dexter, Begelman 2018).

The DRW fit parameters have been linked to the physical quasar properties. Indeed, MacLeod+2010, using the SDSS light curves for Stripe82 found correlations of the characteristic timescale and variability amplitude with the black hole mass, and quasar luminosity.

Quasar variability, and specifically modelling it as a DRW, is also a reliable way to distinguish quasars from stars based on merely optical photometry (MacLeod+2011). In that case the fit biases are less important than the fact that DRW timescale and amplitude for QSO are order of magnitude different from stars (MacLeod+2011). It is especially useful for quasars that could not be easily identified by color-color diagrams (Sesar+2007) - those in the intermediate redshift range, whose colors mimic the M dwarfs (Yang+2017).

Accurate QSO population studies is important for measurement of QLF, and variability has been used before to increase the completeness of quasar selection (Richards+2008, Ross+2013, Palanque-Delabrouille+2013,2016 , AlSayyad+2016, McGreer+2013,2017).

Because DRW is a stochastic process, fitting is more involved than with simpler source variability, such as RR Lyrae, or Eclipsing Binaries that result in well-defined light curve shape. Indeed, even with identical input parameters, two different DRW light curves would have a very different appearance. It has been found (eg. MacLeod et al. (2011) , Kozlowski+2010, Kozłowski, Szymon (2017)) that regardless of method, we can most reliably recover input parameters if we use the longest light curve baseline possible. A rule of thumb is that the light curve has to be at least ten times longer than the recovered timescale. We confirm this observation with simulations of DRW light curves spanning a variety of ratios of input timescale to light curve lenght.

The light curve baseline is the key in an unbiased recovery of light curve parameters. As it has been 8 years since MacLeod+2010 have published their research, we can now benefit from additional data from other surveys that have observed the same quasars since. We show how combining the SDSS data with CRTS, PTF , PS1, and simulated LSST data, decreases the bias in recovered parameters. Thus with added data, extending the baseline by 50% on average, we revisit correlations studied by MacLeod+2010. We confirm the general trends, and provide forecast for improvement with the advent of ZTF, LSST. Extended baseline is the advantage that is not afforded by studies only using single survey data (eg. Hernitschek+2016)

2. METHODS

We use a new Gaussian Process modeling tool, Celerite, for light curve fitting (Foreman-Mackey+2017).

Corresponding author: Krzysztof Suberlak
suberlak@uw.edu

Celerite allows the user to select a variety of kernels describing the Gaussian Process. A kernel is a covariance function - a measure of correlated information content as a function of signal separation. For the DRW process, the kernel is

$$k(\Delta t_{nm}) = a \exp\left(-t_{tm}/\tau\right) = \sigma^2 \exp\left(-t_{tm}/\tau\right) = \sigma^2 ACF(\Delta t) \quad (1)$$

The latter equality shows that kernel is also known as covariance : $S_{DRW}(\Delta t) \equiv k(\Delta t)$, with the amplitude expressed in terms of $\sigma^2 \equiv a$. The last equality shows that for DRW, the correlation is related to the autocorrelation function ACF by variability amplitude .

The structure function SF of the DRW process, which expresses the rms of magnitude differences Δm as a function of temporal separation Δt, is :

$$SF(\Delta t) = SF_\infty (1 - \exp\left(-|\Delta t|/\tau\right))^{1/2} \quad (2)$$

where SF_∞ is the asymptotic value of SF for large time lags (it is known that for QSOs SF follows approximately power law, $SF \propto \Delta t^\beta$ (see MacLeod+2012, etc), and it levels out for large Δt)

Note that $SF_\infty \equiv \sigma^2$ in the above.

(cite MacLeod+2012, Bauer+2009, Graham+2015 for SF overview)

This is equivalent to kernels used by Rybicki+, Kozlowski+, Kochanek+, MacLeod+. We use a prior uniform in log space (Jeff2) - same as MacLeod+, Kochanek+, Kozlowski+.

Given the non-Gaussian shape of the log-posterior, we find that the expectation value rather than the MAP estimate for σ, τ is less biased. Fig.... illustrates the shape of the log Posterior. This was also encountered by Kozlowski+2017 (see his Fig.2).

We first confirm Kozłowski, Szymon (2017) results, simulating DRW light curves with fixed length (baseline) of 8 years, but spanning a grid of timescales so that the ratio of timescales to baseline, $\rho = \tau_{in}/t_{exp} \in \{0.01 : 15\}$. The observed light curve $y(t)$ consists of true underlying signal $s(t)$, noise offset $n(t)$ (which we assume to be drawn from a Gaussian distribution $\mathcal{N}(0, \sigma(t))$ with a width $\sigma(t)$ corresponding to the photometric uncertainty at the given epoch, $e(t)$), and the added mean m :

$$y(t) = s(t) + n(t) + m \quad (3)$$

The underlying DRW signal $s(t)$ is simulated by drawing points from a Gaussian distribution, for which mean and standard deviation are re-calculated at each timestep. Given the array of timestamps to sample the DRW t, variability amplitude SF_∞, mean magnitude m, and the damping timescale τ, we start at time t_0. Signal at that time is equal to the mean $s_0 = m$. After timestep $\Delta t_i = t_{i+1} - t_i$, the signal s_{i+1} is drawn from $\mathcal{N}(loc, stdev)$, where :

$$loc = s_i e^{-r} + m\left(1 - e^{-r}\right) \quad (4)$$

and

$$stdev^2 = 0.5\,\mathrm{SF}_\infty^2\left(1 - e^{-2r}\right) \quad (5)$$

with $r = \Delta t_i/\tau$. This follows eqs. A4 and A5 in Kelly et al. (2009), as well as Sec. 2.2 in MacLeod et al. (2010), and is equivalent to Kozłowski, Szymon (2017) formulation.

In the first simulation setup where we probe a range of ρ ratios, we fixed the baseline to 8 years, variability amplitude $SF_\infty = 0.2$ mag, and used SDSS S82-like cadence with N=60 epochs, or OGLE-III like cadence with N=445 epochs. The errors (in units of mag) were set by the adopted mean magnitudes, $r = 17$ and $I = 18$, as in Kozłowski, Szymon (2017) :

$$\sigma^2_{SDSS} = 0.013^2 + \exp\left(2(r - 23.36)\right) \quad (6)$$

$$\sigma^2_{OGLE} = 0.004^2 + \exp\left(1.63(I - 22.55)\right) \quad (7)$$

with by showing that regardless of cadence, the recovered DRW characteristic timescale becomes independent of the input timescale when it is

Using a fixed light curve length of $t_{exp} = 8$ years we simulate 100 different input time scales τ_{in} , where . The bias is not affected by different cadence (we tested SDSS and CRTS cadences of N=80 or N=445 points). Encouraged by this result we add more epochs to quasar light curves, and revisit relations studied by MacLeod et al. (2011) and Hernitschek et al. (2016). We use PanSTARRS (Chambers 2011) DR2 (Flewelling 2018), CRTS DR2 (Drake et al. 2009), and PTF (Rau et al. 2009). Starting from the SDSS DR7 near-simultaneous ugriz photometry for Stripe 82 quasars (Schneider et al. 2008), we cross-match the catalogs. We find a common photometric solution using the S82 standard stars (PTF and CRTS use white light - see Djorgovski et al. 2011). We validate data quality by comparing mean CRTS magnitude vs mean SDSS magnitude, and mean PS1 g-band to mean PS1 r-band.

3. DATASETS

We focus on data pertaining to a 290 deg^2 region of southern sky, repeatedly observed by the SDSS between 1998 and 2008. Originally aimed at supernova discovery, objects in this area, known as Stripe82 (S82), were re-observed on average 60 times (see MacLeod et al.

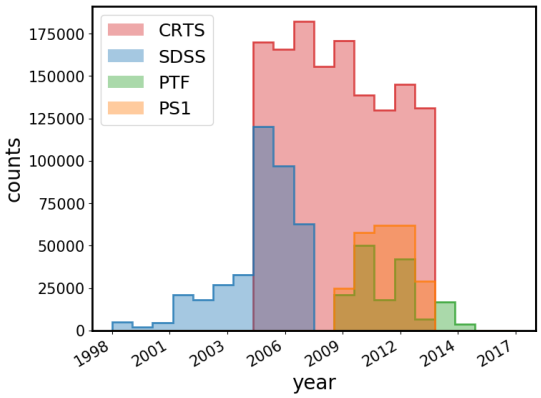

Figure 1. Raw photometric measurements for quasars in Stripe 82 from SDSS(r), PS1(gri), PTF(gR), CRTS(V).

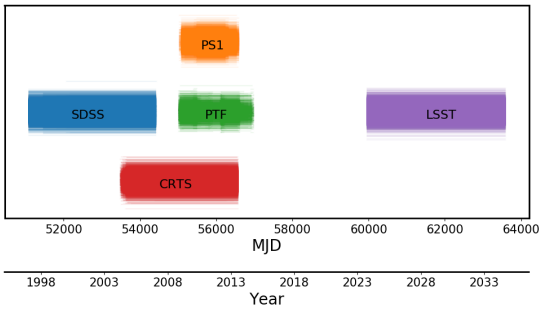

Figure 2. The contribution to quasars light curve baseline from surveys, including the planned LSST coverage. Vertical offset is arbitrary. Note how PS1 and PTF extend the baseline of SDSS by approximately 50%, and how inclusion of LSST triples the SDSS baseline.

2012 Sec. 2.2 for overview, and Annis et al. 2014 for details) . Availability of well-calibrated (Ivezić et al. 2007), long-baseline light curves spurred variability research (see Sesar et al. 2007). The catalog prepared by (Schneider et al. 2008) as part of DR9 contains 9258 spectroscopically confirmed quasars.

We supplement the SDSS with PanSTARRS (PS1) (Chambers 2011; Flewelling 2018), CRTS (Drake et al. 2009), and PTF (Rau et al. 2009). We find 9248 PS1 matches, 6455 PTF matches, and 7737 CRTS matches. Performing an inner join there are 6444 quasars with SDSS-PS1-PTF-CRTS data. Fig. 1 shows the distribution of raw epochs, and Fig 2 the baseline coverage of various surveys. Each survey uses a unique set of bandpasses and cadences : SDSS light curves contain near-simultaneous $\{u, g, r, i, z\}_{SDSS}$, and the other are non-simultaneous : $\{g, r, i, z, y\}_{PS1}$, $\{g, R\}_{PTF}$, V_{CRTS}. To combine the light curves we find color terms with the offset to the 'master' band r_{SDSS}.

Since quasars occupy a blue region in the color-color diagram (Fig. 3), we calculate photometric offsets specifically for this region of the spectrum. We only show the color-magnitude diagrams for PS1 offsets (Fig. 4) since we choose not to include CRTS and PTF data in the final sample.

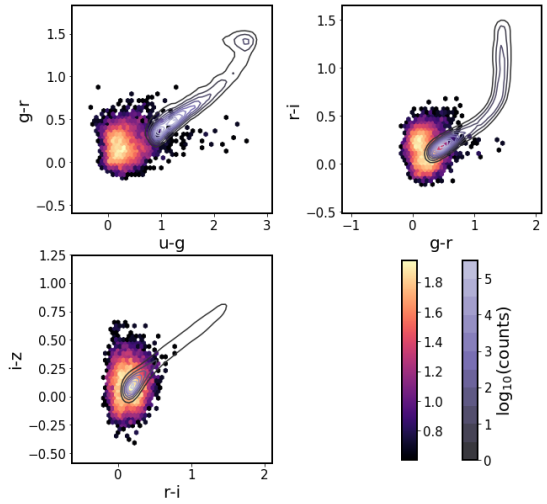

Figure 3. Regions occupied in color-color space by S82 quasars and standard stars (Schneider et al. 2010). We show only 10 000 randomly chosen stars from the full 1 mln + standard stars catalog Ivezić et al. 2007.

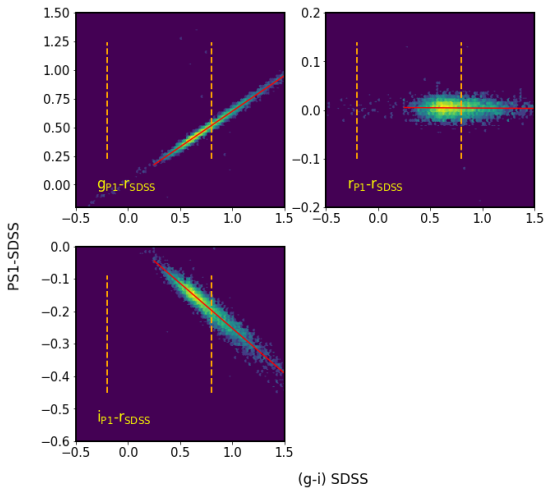

Figure 4. The SDSS-PS1 offsets. We plot only bright stars that have SDSS(r) ¡ 19, and that fulfill mErr * sqrt(Nobs) ¡ 0.03 . Each panel plots about 6000 stars of the 47000 CRTS S82 stars . Vertical dashed lines mark the region in SDSS color space occupied by quasars (see Fig. 3).

With two photometric systems, eg. SDSS(ugriz), and PS1(grizy), we can find offsets (aka 'color terms') from one to another. Consider SDSS as target system, PS1 as the auxiliary system. Thus we are trying to express PS1 in terms of SDSS. This amounts to creating 'synthetic' SDSS bands from PS1, using the SDSS color to spread

the stellar locus. More generally, we would always use the color of the target system.

$$r_{PS1} - r_{SDSS} = f(SDSS(g-i)) \quad (8)$$

the function is a polynomial fitted to the stellar locus on the plot of $SDSS(g-i)$ vs $r_{PS1} - r_{SDSS}$. We use SDSS(g-i) because is provides a larger wavelength baseline. One can make other choices: with $m-s = f(x)$ for instance, Tonry et al. (2012) derived all offsets using x = SDSS$(g-r)$, $m = PS1\{g,r,i,z,y\}$, and s = SDSS(r)

In that way we form a 'master bandpass', consisting of SDSS bands, and PS1, PTF, CRTS equivalents. Since all SDSS bands are observed nearly simultaneously, we choose SDSS(r) as the 'master band', and we transform photometry from all 'nearby' bandpasses in other surveys (PTF gR, PS1 gri, CRTS V) - PS1 u,z,y are too distant from SDSS r. To separate in color space the stellar locus we use SDSS(g-i) color because it has a larger wavelength baseline than SDSS(g-r) color. For each 'master band' light curve we keep track which points originated from SDSS(r), PTF(gR), PS1(gri), or CRTS(V).

Note about EXTINCTION: due to dust present between us and the standard stars (or background quasars), the observed light will appear slightly redder because dust preferentially scatters blue light away. This depends on the location of the source on the sky and is related to the dust inhomogeneities in the Milky Way.

However, in deriving bandpass to bandpass transformation all that matters is the flux received at the Top of the Atmosphere (TOA), for which all photometry is corrected at the pipeline processing level (both for SDSS, PTF, PS1, and CRTS). In other words, all that matters is that any SED at TOA will have slightly different magnitude (hence colors) depending on whether we observe it with SDSS(r), PS1(r), or PTF(R). Therefore, to derive photometric offsets (which are time-independent), we use Sloan colors not corrected for extincton (and likewise, PS1, PTF, etc.) .

Correction for insterstellar extinction is only needed for color selection, since otherwise objects would appear to be of the later stellar type (redder) than they really are. To correct for extinction we use the most up-to-date maps of stellar extinction Bayestar17 (Green+2018). These extinction maps are 3D probabilistic, and we assume a uniform distance of 4 kpc for the dust column for all stars.

3.1. *Simulated light curves*

We made a controlled experiment of long (20 tau) , well-sampled (dt=5 days) light curves, with 400 points each . We used different priors (Jeff1, Jeff2, p1, p2, flat), and found that sigma, tau from MAP (maximum a posteriori estimate) with Jeff1 is most consistent with Chelsea's code. We further investigated the logL evaluated on a grid of sigma,tau, and conclude that the non-Gaussian shape of the log-likelihood cauese the MAP to be biased, and the expectation value of marginalized posterior distribution is less biased (at 1% level) . We find that the expectation value based sigma, tau are less biased, and for a very coarse grid 25x25 elements, the value of input parameters is still recovered at the 1% level (we expect the overall distribution to be biased on the 1% level, so this is sufficient accuracy).

We first check whether there is an improvement of fit for simulated DRW sampled at observed cadence - we plot τ_{out} vs τ_{in} for SDSS sampling, PS1+SDSS sampling, PS1+CRTS+SDSS sampling, PS1+CRTS+PTF+SDSS sampling. This helps establish, based on simulated data (where we know the truth), whether we should expect much improvement in fit accuracy when using real data.

Using the best combination of survey data, we revisit MacLeod et al. (2011) correlations of retrieved characteristic quasar timescale τ and variability amplitude σ with black hole mass, luminosity, etc.

REFERENCES

Annis, J., Soares-Santos, M., Strauss, M. A., et al. 2014, ApJ, 794, 120, doi: 10.1088/0004-637X/794/2/120

Chambers, K. C. 2011, in Bulletin of the American Astronomical Society, Vol. 43, American Astronomical Society Meeting Abstracts #218, 113.01

Djorgovski, S. G., Drake, A. J., Mahabal, A. A., et al. 2011, ArXiv:1102.5004. https://arxiv.org/abs/1102.5004

Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, ApJ, 696, 870, doi: 10.1088/0004-637X/696/1/870

Flewelling, H. 2018, in American Astronomical Society Meeting Abstracts, Vol. 231, American Astronomical Society Meeting Abstracts 231, 436.01

Hernitschek, N., Schlafly, E. F., Sesar, B., et al. 2016, The Astrophysical Journal, 817, 73

Ivezić, Ž., Smith, J. A., Miknaitis, G., et al. 2007, AJ, 134, 973, doi: 10.1086/519976

Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009, The Astrophysical Journal, 698, 895

Kozłowski, Szymon. 2017, A&A, 597, A128, doi: 10.1051/0004-6361/201629890

MacLeod, C. L., Ivezić, Ž., Kochanek, C. S., et al. 2010, The Astrophysical Journal, 721, 1014

MacLeod, C. L., Brooks, K., Ivezić, Ž., et al. 2011, The Astrophysical Journal, 728, 26

MacLeod, C. L., Ivezić, Ž., Sesar, B., et al. 2012, The Astrophysical Journal, 753, 106

Rau, A., Kulkarni, S. R., Law, N. M., et al. 2009, PASP, 121, 1334, doi: 10.1086/605911

Schneider, D. P., Hall, P. B., Richards, G. T., et al. 2008, VizieR Online Data Catalog, 7252

Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, VizieR Online Data Catalog, 7260

Sesar, B., Ivezić, Ž., Lupton, R. H., et al. 2007, AJ, 134, 2236

Tonry, J. L., Stubbs, C. W., Lykke, K. R., et al. 2012, ApJ, 750, 99, doi: 10.1088/0004-637X/750/2/99