Draft version December 17, 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 (DRW) model (Kelly et al. 2009; MacLeod et al. 2010; Kozłowski et al. 2010; Zu et al. 2011; Kasliwal et al. 2015). The origin of variability is debated, with thermal origin being the favorite explanation (Kelly et al. 2013), connected to the inhomogeneity of the accretion disk (Dexter & Agol 2011), or even magnetically elevated disks (Dexter & Begelman 2019).

The DRW parameters have been linked to the physical quasar properties: MacLeod et al. (2010) found correlations of the characteristic timescale and variability amplitude to the black hole mass, and quasar luminosity.

Inherent variability, and modelling it as a DRW, is also a reliable way to distinguish quasars from stars based on optical photometry (MacLeod et al. 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 et al. 2011). It is especially useful for selecting quasars in the intermediate redshift range that could not be easily identified by color-color diagrams (Sesar et al. 2007; Yang et al. 2017).

Accurate QSO population studies are important for measurement of Quasar Luminosity Function, and variability has been used before to increase the completeness of quasar selection (Ross et al. 2013; Palanque-Delabrouille et al. 2013; AlSayyad 2016; McGreer et al. 2013, 2018).

Corresponding author: Krzysztof Suberlak
suberlak@uw.edu

Because DRW is a stochastic process, two light curves with identical DRW parameters will not look identical. However, once can still fit the available data with the DRW model and recover fit parameters. It has been found (eg. MacLeod et al. (2011); Kozłowski et al. (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 et al. (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 on average by 50%, we revisit correlations studied by MacLeod et al. (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 et al. (2016)

2. METHODS

2.1. *DRW as a Gaussian Process*

DRW (OrnsteinUhlenbeck process) can be understood as a member of a class of Gaussian Processes (GP). Each GP is described by a kernel - a covariance function that contains a measure of correlation between two points

x_n, x_m, separted by Δt_{nm}. For the DRW process, the kernel is

$$k(\Delta t_{nm}) = a \exp\left(-\Delta t_{tm}/\tau\right) \quad (1)$$
$$= \sigma^2 \exp\left(-\Delta t_{tm}/\tau\right) \quad (2)$$
$$= \sigma^2 ACF(\Delta t_{tm}) \quad (3)$$

Here a or σ^2 is an amplitude of correlation decay as a function of t_{tm}, while τ is the characteristic timescale over which correlation drops by $1/e$. For a DRW, the correlation function $k(\Delta t_{nm})$ is also related to the autocorrelation function ACF.

Related to $k(\Delta t_{nm})$ is the structure function of the DRW process (see MacLeod et al. (2012); Bauer et al. (2009); Graham et al. (2015) for an overview) , which expresses the rms of magnitude differences Δm as a function of temporal separation Δt, is :

$$SF(\Delta t) = SF_\infty (1 - \exp\left(-|\Delta t|/\tau\right))^{1/2} \quad (4)$$

where SF_∞ is the asymptotic value of SF for large time lags. It is known that for QSOs SF follows approximately power law, $SF \propto \Delta t^\beta$, and it levels out for large Δt) (see MacLeod et al. (2012)). Note that $SF_\infty = \sqrt{2}\sigma$ in the above.

To estimate the DRW parameters and fit simulated or real data we employ Celerite - a new fast GP modelling tool(Foreman-Mackey et al. 2017). Combined with the DRW kernel this is similar to the method used by Rybicki & Press (1992); Kozłowski et al. (2010); MacLeod et al. (2010) - like in previous work, we use a prior on σ and τ uniform in log space. The main difference is that rather than adopting the Maximum A-Posteriori (MAP) as the 'best-fit' value for sought parameters, we find the expectation value of the marginalized log posterior. If the posterior space was a 2D Gaussian in σ, τ space, the expectation value would coincide with the maximum of the log posterior. However, due to non-Gaussian shape of the log posterior, we find that the expectation value is a better estimate of σ and τ rather than MAP.

2.2. *The impact of light curve baseline*

Kozłowski, Szymon (2017) reports that we cannot trust any results of DRW fitting unless the light curve length is at least ten times longer than the characteristic timescale. We confirm these generic trends by repeating Kozłowski, Szymon (2017) simulation setup. We model 10 000 DRW light curves with fixed length (baseline) $t_{exp} = 8$ years, $SF_\infty = 0.2$ mag, but with different input timescales. We parametrize the ratio of timescale to baseline by $\rho = \tau/t_{exp}$. Given that the baselines for all light curves are fixed, by selecting different input τ we probe a logarithmic grid in $\rho \in \{0.01 : 15\}$. For each of 100 distinct values of ρ we perform 100 light curve realizations.

To simulate observational conditions we add to the true underlying signal $s(t)$ a noise offset, $n(t)$. Like Kozłowski, Szymon (2017), we assume $n(t)$ 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, $e(t)$) :

$$y(t) = s(t) + n(t) \quad (5)$$

The $s(t)$ is found by iterating over the array of time steps t. At each step, we draw a point from a Gaussian distribution, for which the mean and standard deviation are re-calculated at each timestep. Starting at t_0, the signal is equal to the mean magnitude, $s_0 = m$. After a timestep $\Delta t_i = t_{i+1} - t_i$, the signal s_{i+1} is drawn from $\mathcal{N}(loc, stdev)$, with :

$$loc = s_i e^{-r} + m\left(1 - e^{-r}\right) \quad (6)$$

and

$$stdev^2 = 0.5\,\mathrm{SF}_\infty^2\left(1 - e^{-2r}\right) \quad (7)$$

where $r = \Delta t_i/\tau$, τ is the damping timescale, SF_∞ is the variability amplitude , and m the mean magnitude. This follows the formalizm in Kelly et al. (2009) (eqs. A4 and A5) as well as in MacLeod et al. (2010) (Sec. 2.2), and is equivalent to the setup of Kozłowski, Szymon (2017).

We adopt SDSS S82-like cadence with N=60 epochs, or OGLE-III like cadence with N=445 epochs. The errors were set by the adopted mean magnitudes, $r = 17$ and $I = 18$, as in Kozłowski, Szymon (2017) :

$$\sigma^2_{SDSS} = 0.013^2 + \exp\left(2(r - 23.36)\right) \quad (8)$$
$$\sigma^2_{OGLE} = 0.004^2 + \exp\left(1.63(I - 22.55)\right) \quad (9)$$

Fig 1 shows that as stipulated in Kozłowski, Szymon (2017), the recovered ρ becomes meaningless ('unconstrained') if the available baseline is not at least ten times longer than the underlying timescale. It also means that by extending the baseline we can move from the biased region to the unbiased regime.

Encouraged by this result, we extend the baselines of quasar light curves, and revisit relations studied by MacLeod et al. (2011) and Hernitschek et al. (2016).

3. DATA

3.1. *Surveys*

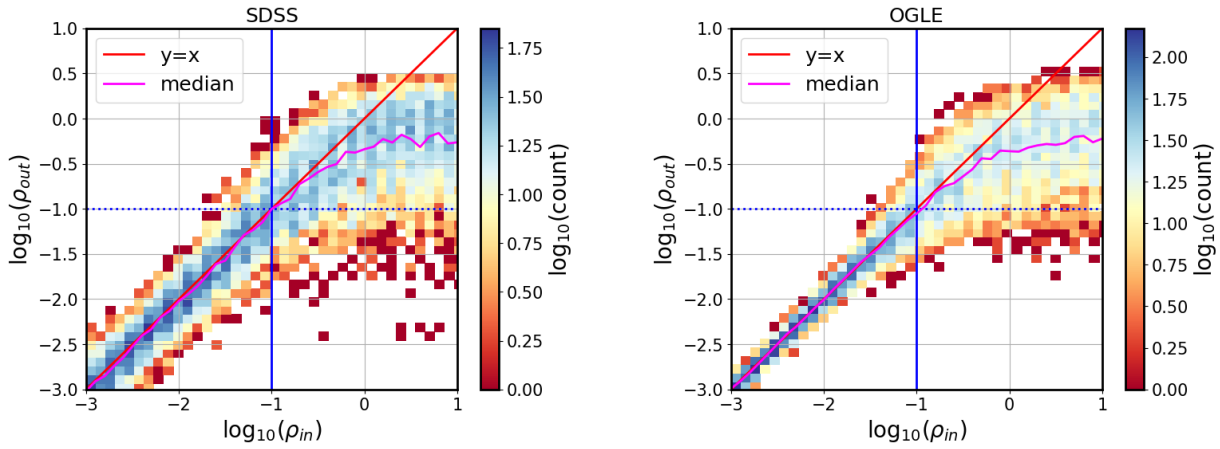

Figure 1. Probing the parameter space of $\rho = \tau/t_{exp}$, with a simulation of 10 000 light curves : 100 light curves per each of 100 ρ values spaced uniformly in logarithmic space between $\rho \in \{0.01 : 15\}$. Thus with the baseline t_{exp} set to 8 years, we sample a range of 100 input timescales, as in Kozłowski, Szymon (2017). Left panel shows the SDSS-like cadence with N=60 points, and the right panel OGLE-like cadence with N=445 points. The dotted horizontal and solid vertical lines represent $\rho = 0.1$, i.e. the baseline is ten times longer than considered timescale. The diagonal line is $y = x$, i.e. the line that would be followed if the recovered ρ (τ) was exactly the same as the input ρ (τ).

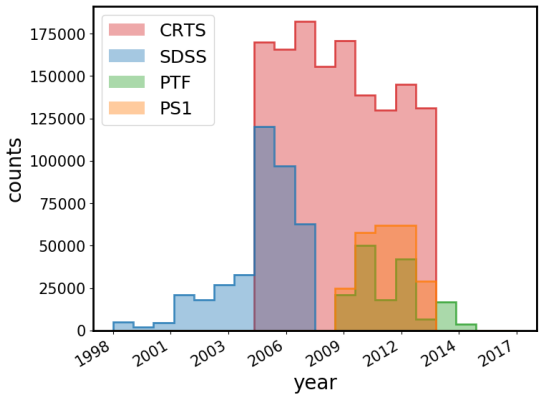

Figure 2. Raw photometric measurements for quasars in Stripe 82 from SDSS(r), PS1(gri), PTF(gR), CRTS(V).

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. 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 extend the SDSS light curves 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 to SDSS S82 quasars. There are 6444 quasars with SDSS-PS1-PTF-CRTS data. Fig. 2 shows the distribution of raw epochs, and Fig 3 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}.

3.2. *Photometric offsets*

To utilize all data we define a common 'target' bandpass. SDSS r band is closest to PS1 r, PTFr, CRTSV. For this reason we translate photometry from nearby filters ($\{g, R\}_{PTF}$, $\{g, r, i\}_{PS1}$, V_{CRTS}) to the 'master' r_{SDSS} band.

With two photometric systems, eg. SDSS(ugriz), and PS1(grizy), we can find offsets (or color terms) from one to another. Consider SDSS as target system, PS1 as the auxiliary system, so that we find offsets from PS1 to 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 :

$$r_{PS1} - r_{SDSS} = f(SDSS(g - i)) \tag{10}$$

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 than $(g - r)$.

Note that there are other possible choices for the target band and the intermediate color to spread the stellar locus. For instance, rewriting the above as $m - s = f(x)$, Tonry et al. (2012) derived offsets from SDSS to PS1

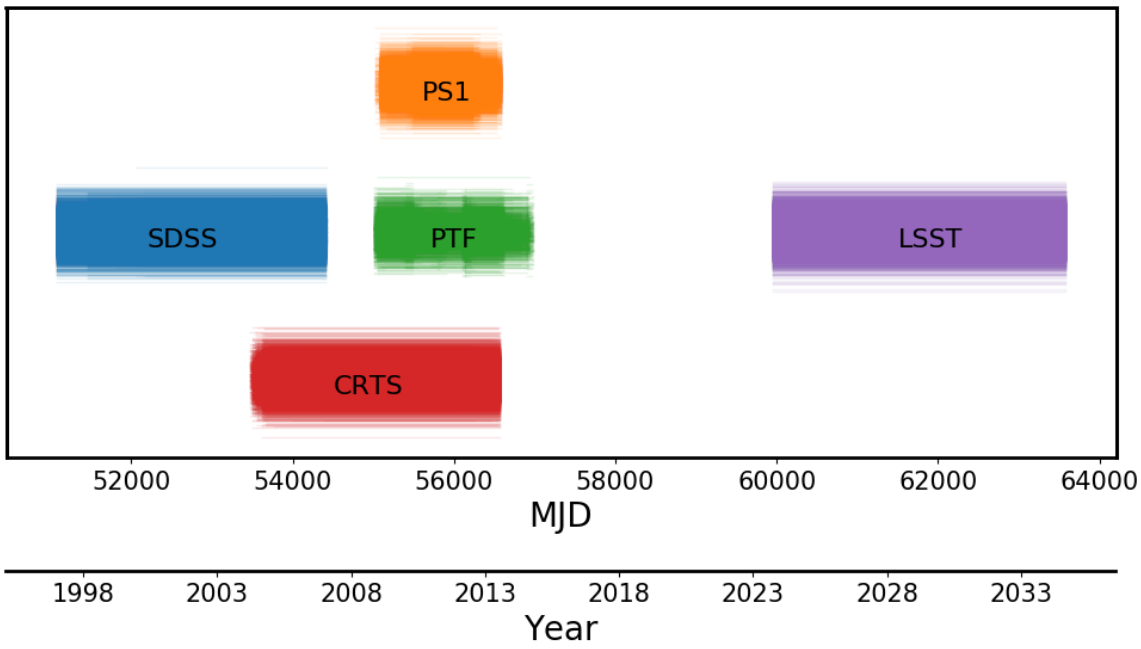

Figure 3. The contribution to quasars light curve baseline from surveys, including the planned LSST coverage. Vertical offset is arbitrary. Note how PS1 and PTF extend the baseline of SDSS by approximately 50%, and how inclusion of LSST triples the SDSS baseline.

using $x = \mathrm{SDSS}(g-r)$, $m = PS1\{g,r,i,z,y\}$, and $s = \mathrm{SDSS}(r)$.

Since quasars occupy a blue region in the color-color diagram (Fig. 4), we calculate photometric offsets specifically for this region of the spectrum. We only show the color-magnitude diagrams for PS1 offsets (Fig. 5) since we choose not to include CRTS and PTF data in the final sample.

4. RESULTS

Having established in Sec. 2.2 that extending the light curve baseline improves the recovery of input DRW parameters, we combine SDSS light curves with PS1 data. We first consider the theoretical improvement in the fit, simulating DRW light curves for which we select SDSS-PS1 , or SDSS-only sections. Then we fit the real data with DRW model, and divide by (1+z) to study the timescales in rest frame.

4.1. *Simulated DRW*

We simulated the DRW using the real cadences and errors corresponding to sections of combined light curves (SDSS,PS1,CRTS,PTF), including the portion corresponding to the forecasted LSST contribution. For LSST segment we assumed a cadence of 50 epochs per year, for 10 years (between 2023-2033), and magnitude-dependent photometric uncertainty from Sec. 3.5 in LSST Science Collaboration et al. (2009):

$$\sigma_{LSST}(m)^2 = \sigma_{sys}^2 + \sigma_{rand}^2 \quad (11)$$

$$\sigma_{rand}^2 = (0.04-\gamma)x + \gamma x^2 \quad (12)$$

$$x = 10^{0.4(m-m_5)} \quad (13)$$

with $\sigma_{sys} = 0.005$, $\gamma = 0.039$, $m_5 = 24.7$ (see Table 3.2 therein).

For all light curves we assumed $\tau = 575$ days, $SF_\infty = 0.2$ mag (the median of S82 distribution in MacLeod et al. (2010)). Fig. 6 shows an example simulated DRW for SDSS-PS1-LSST.

In accordance with Fig. 1, we find that extending the baseline decreases the bias in retrieved DRW parameters τ and σ - see Fig. 7 and Fig. 8

4.2. *Real data*

Since the simulations showed that we do decrease the bias by extending the baseline, we use the combined SDSS-PS1 light curves and show the space occupied by quasars in $SF_\infty - \tau - \sigma$ space, which confirms the results of MacLeod et al. (2011) (Fig. 9)

REFERENCES

AlSayyad, Y. 2016, PhD thesis, University of Washington. http://hdl.handle.net/1773/37020

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

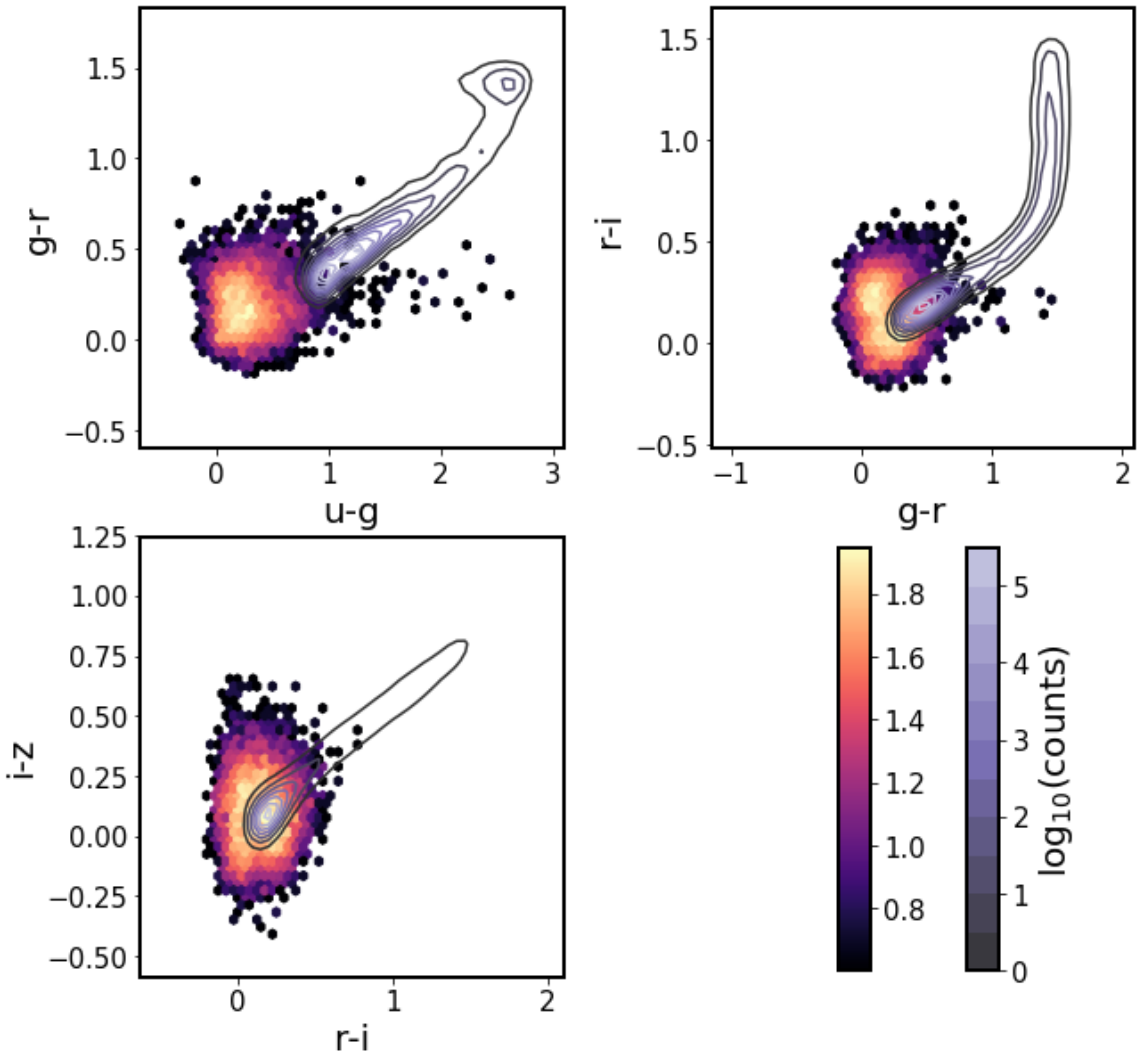

Figure 4. Regions occupied in color-color space by S82 quasars (colors) and standard stars (contours) (Schneider et al. 2010). We show only 10 000 randomly chosen stars from the full 1 mln + standard stars catalog Ivezić et al. 2007.

Bauer, A., Baltay, C., Coppi, P., et al. 2009, ApJ, 696, 1241, doi: 10.1088/0004-637X/696/2/1241

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

Dexter, J., & Agol, E. 2011, ApJL, 727, L24, doi: 10.1088/2041-8205/727/1/L24

Dexter, J., & Begelman, M. C. 2019, MNRAS, 483, L17, doi: 10.1093/mnrasl/sly213

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

Foreman-Mackey, D., Agol, E., Angus, R., & Ambikasaran, S. 2017, ArXiv e-prints. https://arxiv.org/abs/1703.09710

Graham, M. J., Djorgovski, S. G., Stern, D., et al. 2015, Nature, 518, 74, doi: 10.1038/nature14143

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

Kasliwal, V. P., Vogeley, M. S., & Richards, G. T. 2015, MNRAS, 451, 4328, doi: 10.1093/mnras/stv1230

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

Kelly, B. C., Treu, T., Malkan, M., Pancoast, A., & Woo, J.-H. 2013, ApJ, 779, 187, doi: 10.1088/0004-637X/779/2/187

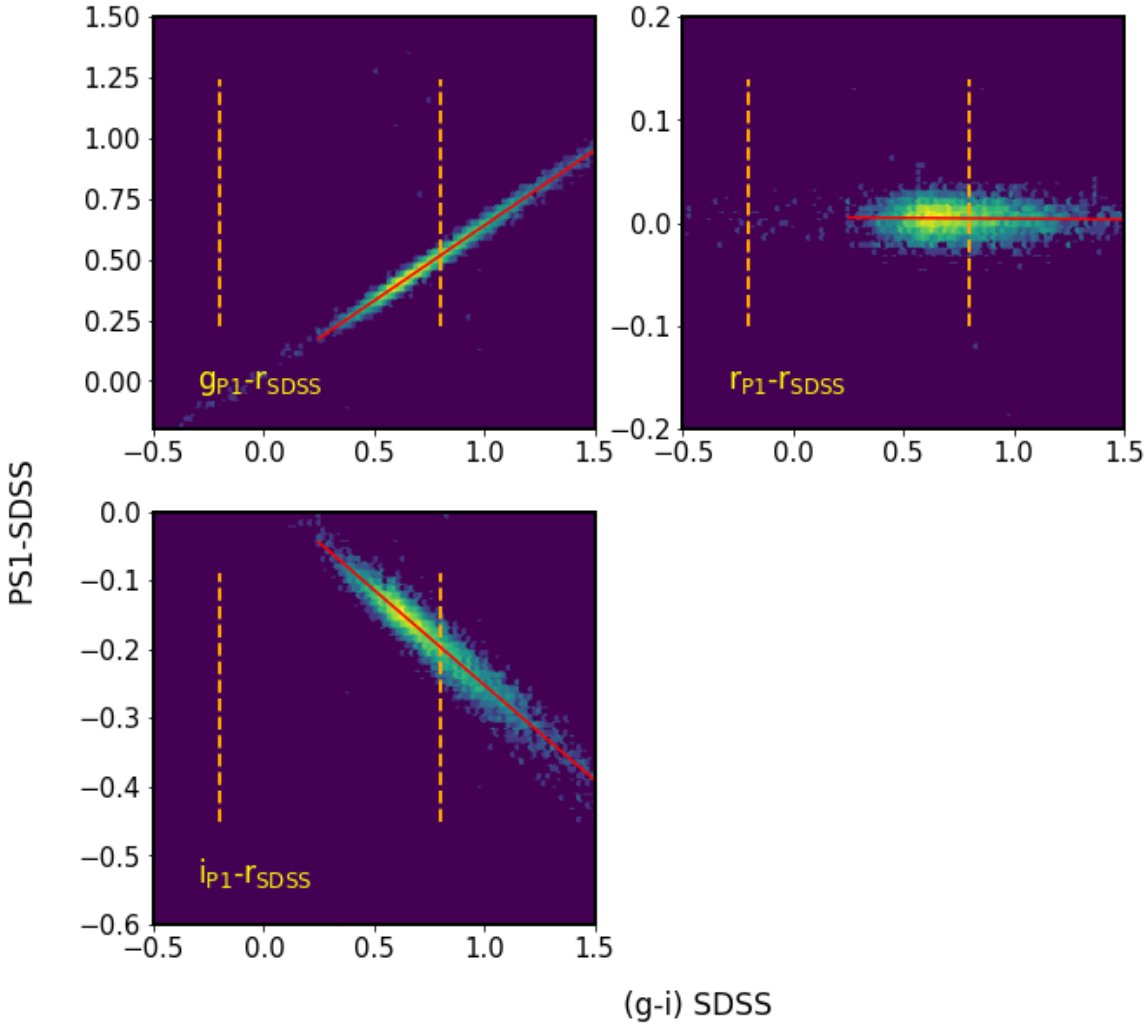

Figure 5. The SDSS-PS1 offsets. We plot only bright stars that have SDSS(r) ¡ 19, and that fulfill mErr * sqrt(Nobs) ¡ 0.03 . Each panel plots about 6000 stars of the 47000 CRTS S82 stars . Vertical dashed lines mark the region in SDSS color space occupied by quasars (see Fig. 4), used to fit the stellar locus with a polynomial.

Kozłowski, S., Kochanek, C. S., Udalski, A., et al. 2010, ApJ, 708, 927

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

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201. https://arxiv.org/abs/0912.0201

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

McGreer, I. D., Fan, X., Jiang, L., & Cai, Z. 2018, AJ, 155, 131, doi: 10.3847/1538-3881/aaaab4

McGreer, I. D., Jiang, L., Fan, X., et al. 2013, ApJ, 768, 105, doi: 10.1088/0004-637X/768/2/105

Palanque-Delabrouille, N., Magneville, C., Yèche, C., et al. 2013, A&A, 551, A29, doi: 10.1051/0004-6361/201220379

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

Ross, N. P., McGreer, I. D., White, M., et al. 2013, ApJ, 773, doi: 10.1088/0004-637X/773/1/14

Rybicki, G. B., & Press, W. H. 1992, ApJ, 398, 169, doi: 10.1086/171845

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

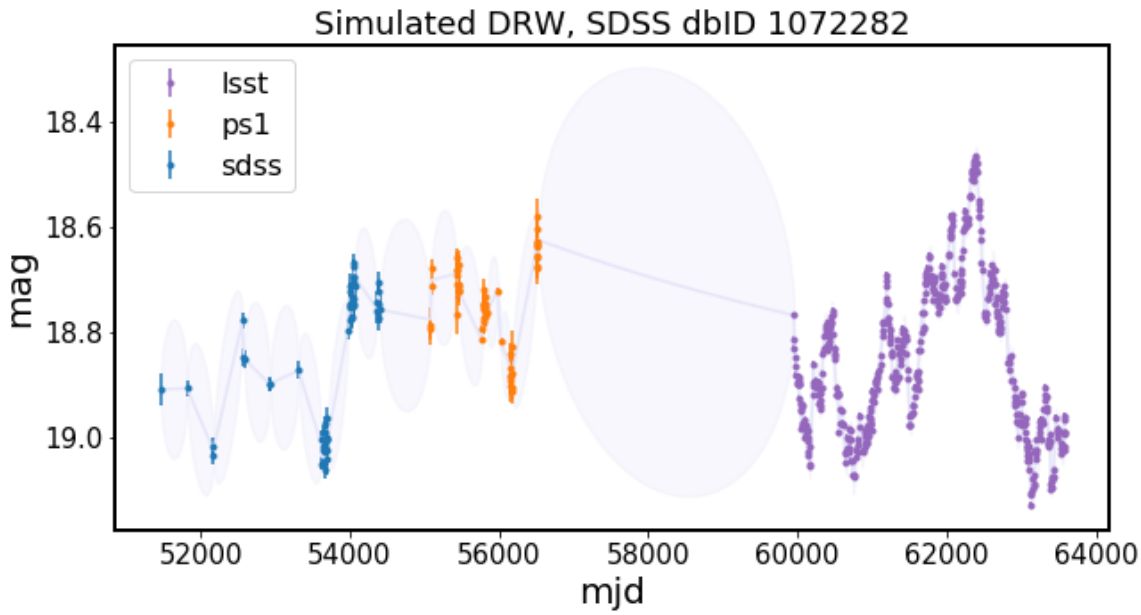

Figure 6. Simulated DRW process sampled at real cadence of SDSS, PS1, and simulated cadence of LSST. To each observed point we add Gaussian noise corresponding to the reported heteroscedastic (different for all points) errors for SDSS-PS1, and simulated magnitude-dependent errors for LSST.

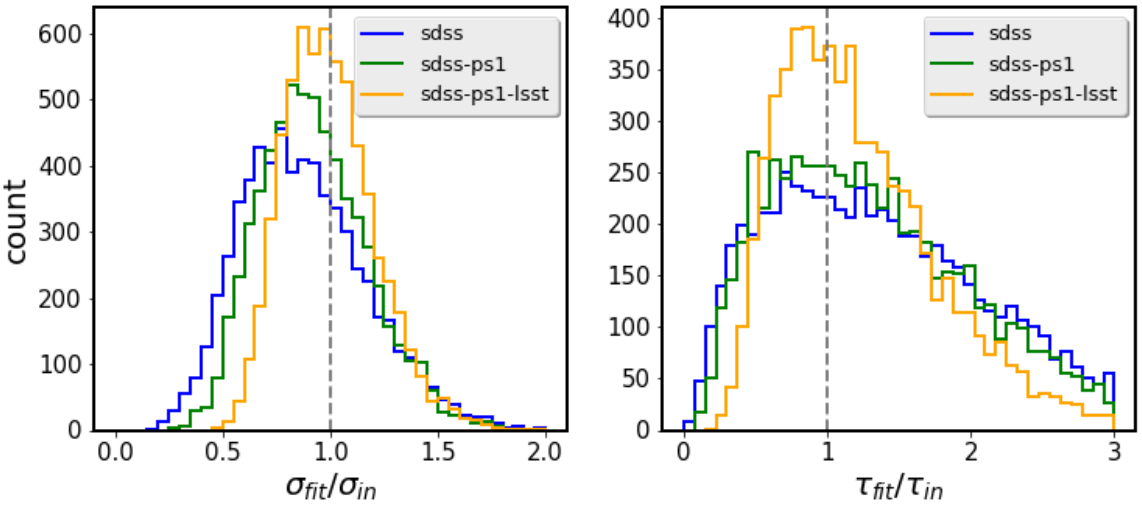

Figure 7. Retrieved τ and σ parameters for simulated LCs.

Yang, J., Fan, X., Wu, X.-B., et al. 2017, AJ, 153, 184, doi: 10.3847/1538-3881/aa6577

Zu, Y., Kochanek, C. S., & Peterson, B. M. 2011, ApJ, 735, 80

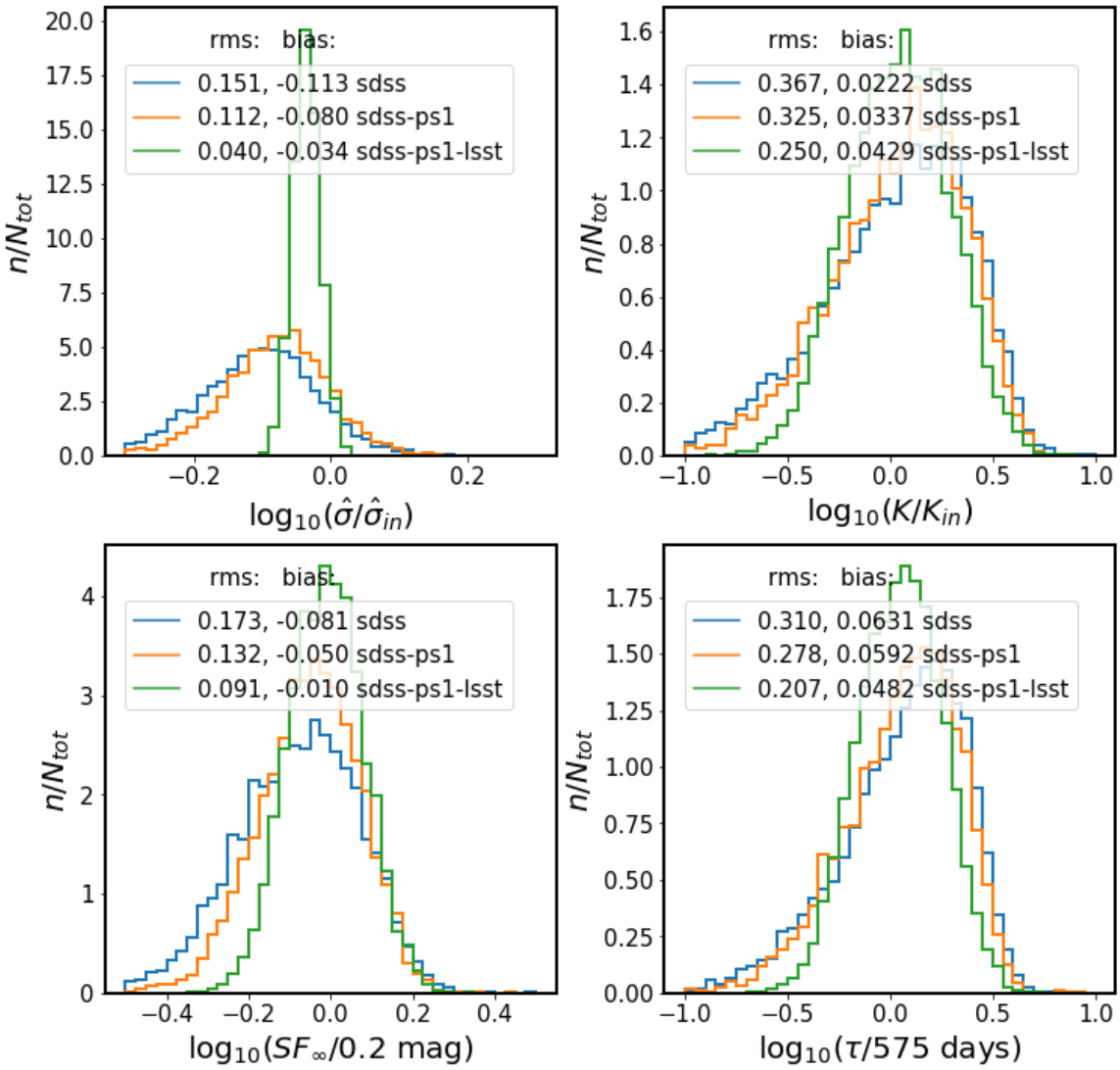

Figure 8. Comparison of retrieved parameters in relation to input parameters, shown as Fig.18 in MacLeod et al. (2011)

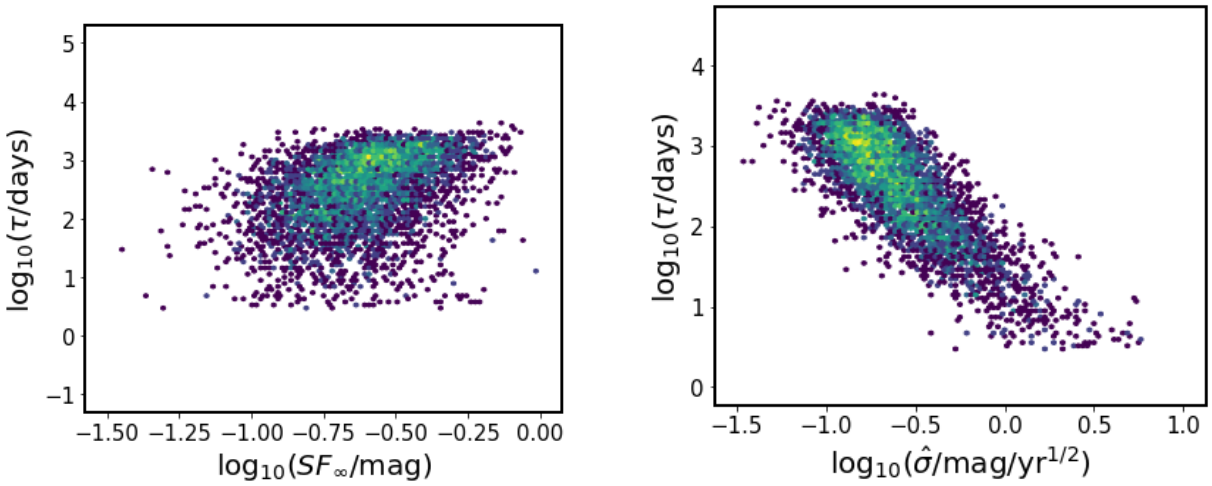

Figure 9. Both relations are shown in the observed frame. The left panel is like Fig.6, and the right like Fig.14 in MacLeod et al. (2011).