

Dissertation Summary  
KRZYSZTOF SUBERLAK<sup>1</sup>  
<sup>1</sup>*University of Washington, Seattle*

(Dated: April 20, 2018)

## ABSTRACT

The dissertation summary....

## PROMPT

”Provide the 10 page dissertation summary to your committee (with a cc: to Christine for departmental records). The summary includes the scientific context for your proposed dissertation, a description of the proposed research and a likely timeline for completion of major milestones.”

## 1. SCIENTIFIC CONTEXT

Quasars are some of the brightest sources of radiation in the Universe. Quasi-Stellar Object is a source of light that appears like an unresolved point source, a star. Yet it does not look like any other star when we look at the entire electromagnetic spectrum, which is very non-thermal, unlike any blackbody.

When observed with a larger telescope, and longer exposure time, we may gather enough light to realize that there is also a faint glow of a host galaxy surrounding it - so much fainter than the quasar light that it was not realized until few decades after a serendipitous discovery of the first quasar. A quasar is an emission from the hot accretion disk surrounding a supermassive black hole in the center of the host galaxy. Such active galactic nuclei - AGN - have been observed in a more local universe, where we can distinguish the intense radiation of jets of matter directed away from the disk, radio-lobes of heated gas emitted away from the central maelstrom, and the host galaxy. Quasars are orders of magnitude brighter than local AGN, and for this reason (and their cosmological distance) have traditionally deserved this separate nomenclature.

Because quasars are present across the observable universe, from the earliest newly-forming galaxies in the recently reionized universe, to those galaxies that have already formed billions of years earlier, they are an excellent probe of how the universe and its content evolved over time. Since quasars are located at cosmological distances (usually redshift  $> 0.1$ , peak of distribution at redshift  $\approx 2$ ), their light passes through other galaxies and clouds of gas before it reaches us. Those leave imprint on the

light of the quasar, and for this reason quasar spectra are used for studies of the intergalactic medium (Prochaska et al. (2014) and references therein), He II reionization history (Khrykin et al. 2017) , Damped Lyman  $\alpha$  Absorbers (Wolfe et al. (2005) for overview, also see more recent Parks et al. (2018) about the effort to find DLA by deep learning of QSO spectra, and Murphy & Bernet (2016) for study of dust content in DLA from SDSS). Quasars shine light on the epoch of reionization - the moment when the universe, full of neutral Hydrogen and Helium from the Big Bang, became ionized. As AlSayyad (2016) points out, the shape of the Quasar Luminosity Function is connected to the SMBH growth, as well as the cosmic reionization. QLF contains information about the amount of quasar contribution in reionizing the universe (Glikman et al. 2011; Masters et al. 2012; Ross et al. 2013). Since quasars accompany a phase of a rapid growth of a supermassive black hole, studying their duty cycle helps properly understand the growth of galaxies, AGN feedback. Quasar luminosity function is directly related to the evolution of the galactic initial mass function (McGreer et al. 2013).

Quasars are also very interesting on their own right. An accretion disk is an extreme environment. It is a disk of viscously heated matter inspiralling onto a supermassive black hole (Ruan 2017). The disk itself is merely a few times bigger than the solar system - (Mudd et al. 2017) uses reverberation mapping to show that for the majority of their sample disk sizes at 2500 angstroms are between 1-10 light days. Compare that to the orbit of pluto, of approximately 0.2277 light days. The disk itself due to such small size (compared to the galaxy itself) cannot be resolved with a telescope. Yet, the emission of the accretion disk overwhelms all the light from the stars of the host galaxy. In case of quasars, the accretion disk is so much brighter than all the stars of the host galaxy, that initially one only sees the light that appears to all come from one point in space - exactly like in case of a star. Hence the light from an accretion disk appears point-like : quasi-stellar. However, if one allows to collect the light for much longer, with bigger collecting area we begin to see the stellar light from the galaxy hosting the active nucleus (Hutchings et al. 2002; Kotilainen et al. 2013; Falomo et al. 2014; Liuzzo et al. 2016; Bayliss et al. 2017).

Although we cannot resolve the accretion disk, it turns out that thanks to its dynamic behavior - the matter actively accreting onto the central supermassive black hole - its brightness varies over time, on a diverse range of timescales (Schawinski et al. 2015). This accretion disk flicker can be linked to the physical properties of the disk : its extent, viscosity, and processes driving its evolution, which helps understand the process of growth of supermassive black holes. Since for a quasar the emission from the accretion disk is dominant compared to the stellar light from the host galaxy, we can assume that all the coherent variability in the quasar light curve is related to the accretion disk variability . Indeed, the discovery that the quasar light can vary significantly over a period of weeks led to implication that the size of the accretion disk cannot be larger than few light weeks, due to causality.

Now it is generally understood that the characteristic timescale of this stochastic variability can be related to the thermal timescales of the disk (Kelly et al. 2007; Zu et al. 2013; Kozłowski et al. 2016). Recent reverberation mapping studies largely confirm this picture (Sun et al. 2015). One way of describing the variation in brightness over time is to plot the signal against time - that is the light curve. A light curve can be further characterized by a structure function, that describes the spread in magnitude differences between pairs of points as a function of time separation. Structure Function tells us how much correlation there is between parts of light curve separated by a given time.

Structure function is also related to the power spectral density. SF can be seen as a mirror image of PSD, in a sense that SF is a function of time differences, while PSD is a function of frequency, which is the inverse of time differences. Processes that have a flat PSD - having an equal power across the entire frequency spectrum - are described as "white" noise, like white light that includes all colors. To further this analogy, stochastic processes that can be described by a simple power law, and yet depart from "white" noise, are termed "colored" noise, depending on the relative amplitude of PSD at various frequencies. Thus we have "red", "pink", or "blue" noise (see Appendix C in Kasliwal et al. (2017) for further details). (Also, great illustration of how a light curve looks like depending on an exponent of a PSD is found in MacLeod et al. (2010), Fig. 9).

Over the last decade we have significantly improved our understanding of quasars, and new variability-based classification techniques (Fan et al. 2001; Richards et al. 2006; Kozłowski et al. 2010; Palanque-Delabrouille et al. 2011; MacLeod et al. 2011; Graham et al. 2014; AlSayyad 2016; Ruan 2017) have yielded an unprecedented number of quasars that allows a statistical study. The new quasar catalogues (eBOSS, etc) allowed to further constrain the Quasar Luminosity Function (Ross et al. 2013; Myers et al. 2015; Palanque-Delabrouille et al. 2016). Classically quasars were selected using color cuts, which mainly targets quasars at redshifts around  $0.5 < z < 2.5$  (Large Bright Quasar Survey, Hewett et al. (1995), the 2dF QSO Redshift Survey Croom et al. (2004), and the 2dF-SDSS LRG and QSO Survey Croom et al. (2009)). In addition to color space, SDSS I,II selected quasars via radio matches to the FIRST survey, or x-ray matches to ROSAT survey - see introduction of (Myers et al. 2015). Since then other classification parameters have been tested, including radio source matching (McGreer et al. 2009), near-IR color cuts (Banerji et al. 2012), radio + near-IR (Glikman et al. 2012), mid-IR (Stern et al. 2005; Richards et al. 2009; Stern et al. 2012), X-rays (Trichas et al. 2012), midIR + X-rays (Lacy et al. 2004; Hickox et al. 2007, 2009), variability (Schmidt et al. 2010; Butler & Bloom 2011; MacLeod et al. 2011; Palanque-Delabrouille et al. 2011). There are some studies that combined color and variability information, eg. Tie et al. (2017); Peters et al. (2015); Sesar et al. (2007).

A better understanding of how well the Damped Random Walk model can be used to model quasar variability is crucial for using DRW model best-fit parameters to characterize and classify quasars. It is a relatively recent field : [Kelly et al. \(2009\)](#) showed that DRW parameters can be linked to the physics of the accretion disk. Since [Kozłowski et al. \(2010\)](#) proposed the DRW model for quasar selection, [MacLeod et al. \(2010\)](#) moved the field of quasar variability studies from a handful of objects to a proper statistical study, fitting over 9 000 quasars in SDSS Stripe 82 with the DRW model. [Kozłowski, Szymon \(2017\)](#) showed that when using the DRW model, the length of light curve is the most important predictor of biases for the best-fit parameters.

His main conclusion is that when the light curve length (baseline) is insufficient( less than 10 times the intrinsic timescale ), then the bias will be larger than a few %. He showed that for a light curve of length similar to the intrinsic timescale, we expect the bias in retrieved timescale of a factor of 3, which we confirm in our independent study. We find with Celerite that a longer baseline can decrease the bias, which implies that the extended light curve can help improve the results of MacLeod+2011.

Despite these advances, there are several crucial questions related to quasar variability that remain open. First, despite general agreement that the DRW is a good first-order description of quasar variability ([Zu et al. \(2011\)](#); [Kozłowski et al. \(2010\)](#)), there have been some inquiries to the fidelity of these descriptions on all timescales [Zu et al. \(2013\)](#); [Kasliwal et al. \(2017\)](#); [Sartori et al. \(2018\)](#). In particular, [Graham et al. \(2014\)](#) reported a detection of characteristic restframe timescale of 54 days, which has not been detected before. We aim to reconcile the CRTS data with the SDSS and PTF data for the same S82 quasars. It was surprising given that other studies more well tuned to short timescales did not report such findings ( especially studies involving Kepler K2 AGN light curves with extremely good cadence - see [Kasliwal et al. \(2015\)](#); [Aranzana et al. \(2018\)](#); [Smith et al. \(2018\)](#)).

Second, although [Kelly et al. \(2011\)](#) showed that the DRW is a good description of the quasar variability, [Kozłowski, Szymon \(2017\)](#) suggested that with the method he employed there are biases in retrieved parameters. We want to evaluate the existing tools for fitting light curves (eg. the implementation of [Rybicki & Press \(1995\)](#) method, JAVELIN of [Zu et al. \(2011\)](#), George, Celerite, etc. ), and find the best tool for future studies given their scalability (speed), memory usage, bias, to be prepared for fitting of billions of light curves in the LSST era.

Finally, we consider whether we can improve the "classical" method of light curve classification by using combination kernel Gaussian Processes. In the standard paradigm, to compare models A and B (eg. sinusoidal oscillation vs DRW), one needs to fit - optimize the cost function for model A and model B, calculate metric describing the 'goodness of fit' (eg.  $\chi^2_{DOF}$ ), and then choose a better fitting model based on that metric. We investigate a Bayesian method of fitting to the data a

combination kernel Gaussian process, where one obtains simultaneously maximum a-posteriori estimates for amplitudes of component kernels.

## 2. PROPOSED RESEARCH

### 2.1. *Discrepant quasar timescales*

To answer the first question, we presented in [Suberlak et al. \(2017\)](#) an improved photometric error analysis for the 7 100 CRTS (Catalina Real-Time Transient Survey) optical light curves for quasars from the SDSS (Sloan Digital Sky Survey) Stripe 82 catalogue. The SDSS imaging survey has provided a time-resolved photometric data set, which greatly improved our understanding of the quasar optical continuum variability. Data for monthly and longer time-scales are consistent with a damped random walk (DRW). Newer data obtained by CRTS provided puzzling evidence for enhanced variability, compared to SDSS results, on monthly time-scales. Quantitatively, SDSS results predicted about 0.06 mag root-mean-square (rms) variability for monthly time-scales, while CRTS data showed about a factor of 2 larger rms, for spectroscopically confirmed SDSS quasars. Our analysis has successfully resolved this discrepancy as due to slightly underestimated photometric uncertainties from the CRTS image processing pipelines. As a result, we found that the correction for observational noise was too small and the implied quasar variability was too large. The CRTS photometric error correction factors, derived from detailed analysis of non-variable SDSS standard stars that were re-observed by CRTS, are about 2030 %, and result in reconciling quasar variability behaviour implied by the CRTS data with earlier SDSS results. An additional analysis based on independent light curve data for the same objects obtained by the Palomar Transient Factory provides further supported for this conclusion. In summary, the quasar variability constraints on weekly and monthly time-scales from SDSS, CRTS and PTF surveys are mutually compatible, as well as consistent with DRW model, as described in ([Suberlak et al. 2017](#)).

### 2.2. *Extending the light curve baseline, using the right tools*

We first confirm the scaling relations by [Kozłowski, Szymon \(2017\)](#) by testing the retrieval of simulated light curve parameters with Celerite . Using a similar MAP-based method we find with Celerite that the same bias in output light curve is found, and is primarily dependent on light curve length in relation to input time scale. Using a fixed light curve length of  $t_{exp} = 8$  years we simulate 100 different input time scales  $\tau_{in}$  , where  $\rho = \tau_{in}/t_{exp} \in 0.01 : 15$ . The bias is not affected by different cadence (we tested SDSS and CRTS cadences of N=80 or N=445 points).

Having established that extending the light curve baseline decreases the bias in retrieved DRW time scales we set out to improve on relations measured by [MacLeod et al. \(2011\)](#); [Hernitschek et al. \(2016\)](#) adding more data points to SDSS and PanSTARRS light curves ([Chambers 2011](#)). We do that by augmenting the existing light curves with additional data from PanSTARRS DR2 ([Flewelling 2018](#)) ,

CRTS (Drake et al. 2009), and PTF (Rau et al. 2009). We start from the SDSS DR7 near-simultaneous ugriz photometry for S82 quasars (Schneider et al. 2008). To use combined datasets we cross-match the catalogs, and find a common photometric solution using the S82 standard stars (PTF and CRTS use white light, Djorgovski et al. (2011)). Querying CRTS DR2 database B.Sesar obtained CRTS white light lightcurves for the S82 DR7 quasars. We obtained the PTF light curves from the PTF IRSA database. C. MacLeod provided the PS1 grizy light curves matched to positions from DR7 catalog. We validate data quality by comparing mean CRTS magnitude vs mean SDSS magnitude, and mean PS1 g-band to mean PS1 r-band.

We first simulate 10 000 DRW light curves at the SDSS cadence, and we test how adding more data (corresponding to PS1, PTF, CRTS datapoints) affects the fidelity of the retrieved to input parameters. Confirming that adding more data at these cadences improves the fit, we fit the real data with the DRW model, and using the spectroscopic information about the quasars (such as black hole mass, luminosity - Kelly et al. (2013)) we reconsider relations tested by MacLeod et al. (2011).

We also test how selecting photometry from only a subset of surveys  $\tau_{PS1}$ ,  $\tau_{(PS1+SDSS)}$ ,  $\tau_{SDSS}$ . We also revisit all relations using only bright quasars with  $\tau_{\langle mag \rangle < 19}$ .

### 2.3. Improving the classical model fitting in the framework of Gaussian processes

Classical model selection involves specifying a loss function that can be used to find best-fitting parameters for a given model. A suitable metric, such as  $\chi^2$ , is used to describe the scatter of data around the model. Choosing a better model involves comparing the values of the loss function. A different approach is to simultaneously find the maximum a-posteriori solution for a model that combines the two. Such possibility is afforded by Gaussian Process with combination kernels, implemented in Celerite (Foreman-Mackey et al. 2017). For the last decade Gaussian Processes have become more well known, employing a class of functions that are characterized by covariance between pairs of points in the dataset (Foreman-Mackey et al. 2017). The combination kernel technique has been successfully employed by Angus et al. (2018) to model quasi-periodic oscillation of star spots on top of the sinusoidally varying signal due to stellar rotation. Success of this approach furthers our interest in employing this methodology to mine the data for a sinusoidal signal superimposed on top of the DRW, such as is the case for the binary SMBH (Charisi et al. 2018).

We first test the combination kernel method, simulating a DRW light curve (parametrized by asymptotic amplitude  $SF_\infty$ , and characteristic timescale  $\tau$ ), with added sinusoidal modulation (parametrized by amplitude  $A$ , period  $P$ ). With input characteristic timescale of  $\tau = 100$  days, and regular sampling every  $dt = 5$  days, we explore regimes from  $A \ll SF_\infty$ , to  $A \sim SF_\infty$ , and from  $P \ll \tau$ , to  $P \gg \tau$ :  $A \in 0.01, 0.1, 0.25, 0.5, 0.75 SF_\infty$ ,  $P \in 0.25, 1, 4\tau$ . We aim to find whether we can distinguish between a pure stochastic signal, and stochastic signal with sinusoidal



modulation. We then test the combination kernel Gaussian process on spectroscopically confirmed SDSS S82 quasars, and finally find the combination kernel amplitudes for all SDSS S82 light curves.

### 3. TIMELINE

### REFERENCES

- AlSaiyad Y., 2016, PhD thesis, University of Washington, <http://hdl.handle.net/1773/37020>
- Angus R., Morton T., Aigrain S., Foreman-Mackey D., Rajpaul V., 2018, *MNRAS*, 474, 2094
- Aranzana E., K rding E., Uttley P., Scaringi S., Bloemen S., 2018, *MNRAS*, 476, 2501
- Banerji M., McMahon R. G., Hewett P. C., Alaghband-Zadeh S., Gonzalez-Solares E., Venemans B. P., Hawthorn M. J., 2012, *MNRAS*, 427, 2275
- Bayliss M. B., et al., 2017, *ApJL*, 845, L14
- Butler N. R., Bloom J. S., 2011, *AJ*, 141, 93
- Chambers K. C., 2011, in American Astronomical Society Meeting Abstracts #218. p. 113.01
- Charisi M., Haiman Z., Schiminovich D., D’Orazio D. J., 2018, *MNRAS*, 476, 4617
- Croom S. M., Smith R. J., Boyle B. J., Shanks T., Miller L., Outram P. J., Loaring N. S., 2004, *MNRAS*, 349, 1397
- Croom S. M., et al., 2009, *MNRAS*, 392, 19
- Djorgovski S. G., et al., 2011, preprint (arXiv:1102.5004),
- Drake A. J., et al., 2009, *ApJ*, 696, 870
- Falomo R., Bettoni D., Karhunen K., Kotilainen J. K., Uslenghi M., 2014, *MNRAS*, 440, 476
- Fan X., et al., 2001, *AJ*, 121, 54
- Flewelling H., 2018, in American Astronomical Society Meeting Abstracts #231. p. 436.01
- Foreman-Mackey D., Agol E., Angus R., Ambikasaran S., 2017, preprint, (arXiv:1703.09710)
- Glikman E., Djorgovski S. G., Stern D., Dey A., Jannuzi B. T., Lee K.-S., 2011, *ApJ*, 728
- Glikman E., et al., 2012, *ApJ*, 757, 51
- Graham M. J., Djorgovski S. G., Drake A. J., Mahabal A. A., Chang M., Stern D., Donalek C., Glikman E., 2014, *MNRAS*, 439, 703
- Hernitschek N., et al., 2016, *The Astrophysical Journal*, 817, 73
- Hewett P. C., Foltz C. B., Chaffee F. H., 1995, *AJ*, 109, 1498
- Hickox R. C., et al., 2007, *ApJ*, 671, 1365
- Hickox R. C., et al., 2009, *ApJ*, 696, 891
- Hutchings J. B., Frenette D., Hanisch R., Mo J., Dumont P. J., Redding D. C., Neff S. G., 2002, *AJ*, 123, 2936
- Kasliwal V. P., Vogeley M. S., Richards G. T., 2015, *MNRAS*, 451, 4328
- Kasliwal V. P., Vogeley M. S., Richards G. T., 2017, *MNRAS*, 470, 3027
- Kelly B. C., Bechtold J., Siemiginowska A., Aldcroft T., Sobolewska M., 2007, *ApJ*, 657, 116
- Kelly B. C., Bechtold J., Siemiginowska A., 2009, *ApJ*, 698, 895
- Kelly B. C., Sobolewska M., Siemiginowska A., 2011, *ApJ*, 730, 52
- Kelly B. C., Treu T., Malkan M., Pancoast A., Woo J.-H., 2013, *ApJ*, 779, 187
- Khrykin I. S., Hennawi J. F., McQuinn M., 2017, *ApJ*, 838, 96
- Kotilainen J., Falomo R., Bettoni D., Karhunen K., Uslenghi M., 2013, preprint, (arXiv:1302.1366)
- Kozłowski, Szymon 2017, *A&A*, 597, A128
- Kozłowski S., et al., 2010, *ApJ*, 708, 927

- Kozłowski S., Kochanek C. S., Ashby M. L. N., Assef R. J., Brodwin M., Eisenhardt P. R., Jannuzi B. T., Stern D., 2016, *ApJ*, **817**, 119
- Lacy M., et al., 2004, *ApJS*, **154**, 166
- Liuzzo E., et al., 2016, *AJ*, **152**, 38
- MacLeod C. L., et al., 2010, *ApJ*, **721**, 1014
- MacLeod C. L., et al., 2011, *ApJ*, **728**, 26
- Masters D., et al., 2012, *ApJ*, **755**
- McGreer I. D., Helfand D. J., White R. L., 2009, *AJ*, **138**, 1925
- McGreer I. D., et al., 2013, *ApJ*, **768**, 105
- Mudd D., et al., 2017, preprint, ([arXiv:1711.11588](https://arxiv.org/abs/1711.11588))
- Murphy M. T., Bernet M. L., 2016, *MNRAS*, **455**, 1043
- Myers A. D., et al., 2015, *ApJS*, **221**, 27
- Palanque-Delabrouille N., et al., 2011, *A&A*, **530**, A122
- Palanque-Delabrouille N., et al., 2016, *A&A*, **587**, A41
- Parks D., Prochaska J. X., Dong S., Cai Z., 2018, *MNRAS*, **476**, 1151
- Peters C. M., et al., 2015, *ApJ*, **811**, 95
- Prochaska J. X., Lau M. W., Hennawi J. F., 2014, *ApJ*, **796**, 140
- Rau A., et al., 2009, *PASP*, **121**, 1334
- Richards G. T., et al., 2006, *AJ*, **131**, 2766
- Richards G. T., et al., 2009, *ApJS*, **180**, 67
- Ross N. P., et al., 2013, *ApJ*, **773**
- Ruan J. J., 2017, PhD thesis, University of Washington, <http://hdl.handle.net/1773/40473>
- Rybicki G. B., Press W. H., 1995, *Physical Review Letters*, **74**, 1060
- Sartori L. F., Schawinski K., Trakhtenbrot B., Caplar N., Treister E., Koss M. J., Urry C. M., Zhang C. E., 2018, *MNRAS*, **476**, L34
- Schawinski K., Koss M., Berney S., Sartori L. F., 2015, *MNRAS*, **451**, 2517
- Schmidt K. B., Marshall P. J., Rix H.-W., Jester S., Hennawi J. F., Dobler G., 2010, *ApJ*, **714**, 1194
- Schneider D. P., et al., 2008, VizieR Online Data Catalog, **7252**
- Sesar B., et al., 2007, *The Astronomical Journal*, **134**, 2236
- Smith K. L., Mushotzky R. F., Boyd P. T., Malkan M., Howell S. B., Gelino D. M., 2018, preprint, ([arXiv:1803.06436](https://arxiv.org/abs/1803.06436))
- Stern D., et al., 2005, *ApJ*, **631**, 163
- Stern D., et al., 2012, *ApJ*, **753**, 30
- Suberlak K., Ivezić Ž., MacLeod C. L., Graham M., Sesar B., 2017, *MNRAS*, **472**, 4870
- Sun M., et al., 2015, *ApJ*, **811**, 42
- Tie S. S., et al., 2017, *AJ*, **153**, 107
- Trichas M., et al., 2012, *ApJS*, **200**, 17
- Wolfe A. M., Gawiser E., Prochaska J. X., 2005, *ARA&A*, **43**, 861
- 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