

On possible proxies of AGN light curves cadence selection in future time domain surveys

Andjelka B. Kovačević¹, Dragana Ilić¹, Luka Č. Popović^{1,2}, Viktor Radović¹, Isidora Jankov¹, Ilsang Yoon³, Neven Caplar⁴, Iva Čvorović-Hajdinjak¹, Saša Simić⁵

¹ Department of astronomy, Faculty of mathematics, University of Belgrade, Studentski trg 16, 11000 Belgrade, Serbia, e-mail: andjelka@matf.bg.ac.rs

² Astronomical Observatory, Volgina 7, 11000 Belgrade, Serbia e-mail: lpopovic@aob.rs

³ The National Radio Astronomy Observatory, 520 Edgemont Road Charlottesville, VA 2290, USA e-mail: iyoona@nrao.edu

⁴ Astrophysical Department, Princeton University, 4 Ivy Lane, Princeton, NJ, USA e-mail: ncaplar@princeton.edu

⁵ Faculty of Science, University of Kragujevac, Radoja Domanovića 12, 34000 Kragujevac, Serbia e-mail: ssimic@kg.ac.rs

Received December 20, 2020; accepted ...

ABSTRACT

Aims. Motivated by upcoming photometric and spectroscopic surveys (Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST), Manuakea Spectroscopic Explorer), we design the statistical proxies to measure the cadence effects on active galactic nuclei (AGN) variability-observables (time-lags, periodicity, and structure-function (SF)).

Methods. We constructed a multiple-regression model to statistically identify the cadence-formal error pattern knowing AGN time-lags and periodicity from different surveys. We defined the simple metric for the SF's properties, accounting for the 'observed' SF's deviation relative to those obtained from the homogeneously-sampled light curves. We tested the regression models on different observing strategies: the optical dataset of long light-curves of eight type -1 AGN with certain peculiarities and the artificial datasets based on several idealized and LSST-like cadences. The SFs metric is assessed on synthetic datasets.

Results. The regression models (for real and LSST-like data) predict similar cadences for time-lags and oscillation detection, whereas for light curves with low variability ($\sim 10\%$), cadences for oscillation detection differ. For higher variability ($\sim 20\%$), predicted cadences are larger than for $F_{var} \sim 10\%$. The predicted cadences for time-lag and periodicity are decreasing with redshift. SFs with dense and homogenous cadences are more likely to behave similarly. SFs with oscillatory signals are sensitive to the cadences, possibly impacting LSST-like operation strategy. The estimated probability density of binary-AGN candidates for the LSST-like survey gives general cadence constraints compatible with model-inferred.

Conclusions. The proposed proxies can help to select spectroscopic and photometric-surveys cadence strategies, and they will be tested further in larger samples of objects.

1. Introduction

The extreme limits of physical conditions in the universe are set by an event horizon of a supermassive black hole (SMBH) in the centers of an active galactic nuclei (AGNs). Particularly, information of the size and structure of broad line region (BLR) surrounding SMBH comes mainly from the optical reverberation mapping (RM) campaigns (Peterson et al. 1998, 2004; Shapovalova et al. 2001, 2004, 2008, 2010a,b, 2012, 2013, 2016, 2017, 2019; Bentz et al. 2008, 2009a; Denney et al. 2009; Barth et al. 2011, 2013, 2015; Popović et al. 2011, 2014; Grier et al. 2012; Wang et al. 2014; Du et al. 2014, 2015, 2016, 2018; Shen et al. 2016; Grier et al. 2017; Edelson et al. 2019; Wang et al. 2020). Optical RM has been used for the mass determination of SMBH in broad emission-line AGNs, so-called type 1 AGNs. In contrast to studies that consider the nearby quiescent galaxies, for most AGNs spatially resolved observations are not possible. However, RM allows a time-domain investigation of the SMBH influence on its surroundings through the spectroscopic monitoring of the variability of continuum and the lagged response -reverberation - of broad emission lines. Most of the data suffer from 'Static Illusion', because what we can observe may change on timescales larger than centuries ($10^4 - 10^7$ yr, Elvis 2001).

The problem is more apparent in the periodicity detection in the light curves of AGNs. Most AGN light curves are consistent with random (aperiodic, stochastic, noise) variations called 'red

noise' caused by accretion processes which can mimic periodic behavior (see e.g. Vaughan et al. 2016, and reference therein) which makes periodic oscillations more difficult to detect with classical statistical methods (Vaughan 2010). Most AGNs exhibit up to 0.2 mag rms stochastic variability in the optical band, with a long tail of extreme variability objects (>0.5 mag rms), so-called changing-look AGNs (CLAGN, see e.g. Lyutyj et al. 1984; Kollatschny & Fricke 1985; Denney et al. 2014; Wang et al. 2018; Shapovalova et al. 2019; Ilić et al. 2020). The RM campaigns have shown that for a higher accuracy of time-lag measurement, an optimal combination of the following parameters are needed: i) higher temporal resolution (i.e. cadences), ii) longer duration of light curve, iii) higher signal-to-noise (i.e. better flux measurement accuracy), and iv) higher level of variability (Kim et al. 2019). Precise specification of required sampling rate depends on physical time scales of the source (e.g. minimum and maximum timescales of interest, such as the BLR light-crossing time) and the character of variations (Horne et al. 2004), which in most cases resemble red noise. High accuracy of flux measurement allows us to better examine the level of variability. We expect an improved accuracy of time-lag measurement for objects with larger variability. Finally, the number of data points for a given time lag is important, as denser light curve sampling improves the prominence of light curve features such as 'peaks' and 'valleys', and may help understand any pos-

sible trend in the light curve. Thus cadence can be probed by a statistical metric which will correlate with measured time-lag errors (Kim et al. 2019). Similarly, the detection of possible periodic oscillations in the RM light curves would be improved with denser sampling rates, higher flux measurement accuracy, larger variability (Bon et al. 2012; Graham et al. 2015; D’Orazio et al. 2015; Jun et al. 2015; Li et al. 2016; Bon et al. 2016; D’Orazio & Haiman 2017; Charisi et al. 2018; Kovačević et al. 2017, 2018, 2019, 2020). As in case of time-lags, a metric based on the measured light-curve properties could be used to estimate cadences necessary for the detection of periodic oscillations.

Motivated by upcoming photometric surveys over broad range of wavelengths which will periodically observe a large fraction of the sky with good photometric accuracy (e.g., the Vera C. Rubin Observatory Legacy Survey of Space and Time, LSST), Chelouche & Daniel (2012) show that photometric reverberation mapping of the BLR is feasible. They prove that the different variability properties of continuum and line processes can be separated at the light curve level. Hence, in this work, our empirical analysis of different LSST cadences builds on this photometric reverberation mapping proof-of-concept. However, the statistical analysis of the photometric RM data is not limited to the LSST survey and can be adopted easily by other surveys for example, the forthcoming Manuakea Spectroscopic Explore (MSE). The MSE will expand the observed space in both depth, time and wavelength (The MSE Science Team et al. 2019). Apart from having a 11.25m telescope aperture, another special feature of the design is the state-of-art multi-object-spectrograph that will cover also the infrared spectral band, and thus be able to provide simultaneous coverage of C IV and H β for quasars up to ~ 2.5 . This would be particularly interesting for the reverberation mapping of quasar, since the MSE will significantly improve the depth and redshift range of H β line coverage (Shen et al. 2019). It would be important to assess the cadence strategies for the MSE reverberation campaign of about ~ 5000 quasars, which will be spanned over a period of several years. Therefore, to prepare the operation and maximize the science output, for the future large scale time domain surveys, we need to develop universal statistical proxies for evaluation of different cadence strategies.

In order to contribute towards quantification of the products observing strategies in the era of large synoptic time domain surveys such as LSST, we examine the relationship between the AGN variability related observables (time lag, periodic oscillations and structure function) and the characteristics of the ‘LSST-like’ observations and survey cadences. The selection of cadences in different observing strategies of the LSST-like surveys will affect the accuracy of time-lag estimates from AGN light curves, as well as the detection of periodic oscillations. The latter defines our ability to detect candidates of Close Binary Super Massive Black Holes (CB-SMBH) which will be the targets of the next generation of nano-Hertz gravitational observatories (Burke-Spolaor et al. 2019). With the anticipated number of millions of AGNs that will be observed and newly discovered by LSST, it is not viable to spectroscopically follow up of all of them. Thus, being able to photometrically detect possible periodic signals will allow us to harness the power of the datasets LSST will provide and further contribute to the multi-messenger astronomy. One of the commonly used approaches for AGN variability study is structure function (SF) analysis (see e.g., MacLeod et al. 2010; Caplar et al. 2017, and references therein), firstly introduced by Simonetti et al. (1985). Therefore, in this work we aim to show impacts and associated relation of different observing cadence, both from the previous

realistic observations and the proposed observing cadence from the LSST Operations Simulator (OpSim) outputs (Jones et al. 2020), on detection of AGN variability related observables. For this analysis, we employ two types of data: a compilation of the data from 2-3 decades long AGN monitoring campaigns; and a suit of artificial light curves. The uniform set of the light curves of type 1 AGN was collected during very long RM campaigns (up to 3 decades) presented in Shapovalova et al. (2001, 2004, 2008, 2010a,b, 2012, 2013, 2016, 2017, 2019). The artificial light curve datasets comprises two subsets: simulations with a) idealized 1 day uniform observing cadence and b) LSST OpSim outputs. The idealized artificial data sets are generated based on Damped Random Walk (DRW, Kelly et al. 2009). For testing oscillation detection the certain periods are introduced in light curve simulations. Their cadences correspond to several idealized observing strategies. Also, in similar manner we constructed light curves for different runs from the LSST OpSim outputs (Jones et al. 2020). By constructing proxy variables and then applying them to these two sets of light curves, one obtains statistical features that allow us to predict the suitability of different cadences for time-lag estimations, for very delicate detection of oscillation in light curves, as well as for reproducing the SF properties. The main goal of this study is to put constraints on the cadence of the LSST and similar surveys required to achieve a requisite level of time lag uncertainty and periodicity detection, and reconstruction of SF properties.

To find these constraints, we first develop a suite of statistics (i.e. metrics) to quantify an efficacy of the LSST observation with different cadences (or sampling rates) for estimation of time-lags, detection of CB-SMBH candidates, and SF properties. These are all explained in detail in Section 2, in which we also describe the used data, i.e. the compilation of observed very long AGN light curves, and the generation of artificial set of 10yr-long light curves. Then, in Section 3 we calculate the statistics and characteristics of detectable time-lags and oscillatory signals, and give estimates for their error predictions usable for future time domain surveys. In Section 4, we test the proposed metric for AGN light curve SF on the set of artificial light curves with various observing cadences. We compare different observing cadences in the frame of oscillatory detection and discuss the impact of observing cadence on the variability detection. We summarize the results and give our conclusion in Section 5.

2. Method and data

Large time-domain surveys aim to carefully design their observing strategies in order to meet the requirements of most science cases. As an example, the specific observing strategy that LSST will follow is not completely decided and might not be sufficient to fully probe all AGN variability of interest, e.g. high magnification events (Neira et al. 2020). These potential strategies are (LSST Science Collaborations et al. 2017; Brandt et al. 2018; Jones et al. 2020):

- uniform in both cadence and filters within the high level constraints. This would emphasize detection of longer time scale brightness fluctuation events that can be followed in all LSST bands.
- with “rolling” cadences, intended to follow up shorter time scale events. This assumes emphasizing specific areas of the sky (and observing bands) for a certain period of the survey and then targeting (or “rolling to”) a different area (and/or band).

Therefore, it is important to define metrics, which can be used to characterize the performance of different observing strategies and assess a quality of the detection of oscillatory signal, as described.

In order to compare different observing strategies applicable to LSST, we: (i) construct an ensemble of metrics, (ii) compile three decades long light curves from real monitoring campaigns, and (iii) generate synthetic data points assuming both hypothetical observing strategy, as well as the ones that are observed by LSST. In this section we present the details of the method used for calculating the cadence metrics, and the data used to obtain these empirical metrics.

2.1. Metrics based on statistical proxies of time lag and periodicities uncertainties

There are two important aspects in the AGN light curve that affect the error of the rest-frame time lag (τ) and period of underlying oscillation (P) in the light curve:

- (i) the amplitude of flux variation F_{var} relative to the measured flux (or photometric) error σ : $\frac{F_{\text{var}}}{\sigma}$;
- (ii) the ratio of observed time scale $\mathcal{T}_{\text{obs}} \in (\tau, P)$ and light curve sampling time Δt : $\frac{\mathcal{T}_{\text{obs}}}{\Delta t}$.

In general, one can expect that the error of the time lag or oscillation decreases with increasing $\frac{F_{\text{var}}}{\sigma}$ and $\frac{\mathcal{T}_{\text{obs}}}{\Delta t}$.

Then, by assuming that there is no correlation between the amplitude of flux variation and the measured time scale \mathcal{T}_{obs} , we propose a proxy, $\phi_{\mathcal{T}}$ for the error of measured quantity, \mathcal{T} as follows:

$$\log \phi_{\mathcal{T}} = \log \frac{\sigma_{\mathcal{T}}}{\mathcal{T}} \propto A + C_1 \frac{F_{\text{var}}}{\sigma} + C_2 \frac{\mathcal{T}_{\text{obs}}}{(1+z)\Delta t}, \quad (1)$$

where we assume that the error of the time lag or periodicity inferred from the observed light curve will be increasing with increasing redshift of the object for the same observed time scale and sampling time. This proxy can weight the impact of the contrast of flux variation

$$\frac{F_{\text{var}}}{\sigma}$$

and contrast between observed time scale expressed in the rest frame and the sampling frequency

$$\frac{\mathcal{T}_{\text{obs}}}{(1+z)\Delta t}$$

separately, so the multiple linear regression on the ensemble of light curves using this proxy may give a description of the error of the variable of interest.

One can measure the time lag and period, and their uncertainties for each light curve with different F_{var} and τ or P , and perform multiple linear regression model on the ensemble of measured uncertainties using the designed proxy with

$$\frac{F_{\text{var}}}{\sigma}$$

and

$$\frac{\mathcal{T}_{\text{obs}}}{(1+z)\Delta t}$$

as independent and uncertainties as dependent variable. The model will provide the coefficient A, C_1, C_2 of the error proxy. If errors of time lag or periodicity obtained from RM campaigns correlate with the proxy $\psi_{\mathcal{T}}$, as defined in Eq. 1 then that relationship can be used to predict the minimum temporal sampling (cadence) required to recover the measured quantity within specified accuracy.

2.2. Structure function metric

We adopt the first-order SF method (see discussion in Kozłowski 2016, and references therein), defined as

$$SF(\Delta t) = \sqrt{\frac{1}{N_{\Delta t \text{ pairs}}} \sum_1^{N_{\Delta t \text{ pairs}}} (y(t) - y(t + \Delta t))^2}, \quad (2)$$

where collection of measured data $y = \{y_i\}$, $i = 1, n$ (e.g., magnitudes) at times $t = \{t_i\}$, $i = 1, n$ with $\Delta t = |t_{i+1} - t_i|$ and $N_{\Delta t \text{ pairs}}$ is a number of data pairs with time lag Δt . The idea of a potential metric is to estimate deviations between the SFs for densely and uniformly sampled light curve (SF_{conti}) and those with gaps or variable cadence, termed "gappy" light curve (SF_{gappy}).

Assuming two SF curves, both of which constructed with the same bins of time lag, we can define the deviation of SF as follows,

$$M = SF_{\text{conti}} - SF_{\text{gappy}}. \quad (3)$$

Thus M will be the curve representing the deviations of the SF_{gappy} (based on the gaped light curve) from the SF_{conti} . We can also define the metric for an ensemble of k simulated simulated light curves providing SF_{conti}^i and SF_{gappy}^i , $i = 1, k$ where k is the number of simulated light curves. Then the metric given by Eq.3 becomes:

$$M^i = SF_{\text{conti}}^i - SF_{\text{gappy}}^i, \quad (4)$$

where M^i is the ensemble of the deviations curves of SF. If we calculate these deviations curves for different redshift bins we can average the deviations curves for each redshift bin z .

$$M^z = \frac{1}{N_z} \sum SF_{\text{conti}}^i - SF_{\text{gappy}}^i = \frac{1}{N_z} \sum_{i=1}^{N_z} M^i, \quad (5)$$

where M^z is an averaged deviation curve for redshift bin z , N_z is the number of deviations curves M^i within the redshift bin z . We can plot this averaged SF deviations curves on the redshift and characteristics time scale domain. We note that one can use other metrics based on machine learning methods to measure similarity between two curves as e.g., Machalonobis, Minkowski, cross correlation, etc.

2.3. Data

We used two type of data sets to test our metrics: compiled observations of three decade long light curves and artificially generated light curves.

2.3.1. Compiled data set of observed light curves

We select a sample of objects to be consistent in the sense of instruments used, length of monitoring campaigns, calibration and flux measurements, as well as statistical tools used to infer time lag and periodicity. We use the optical RM data from very long monitoring campaigns of Shapovalova et al. (2001, 2004, 2008, 2010a,b, 2012, 2013, 2016, 2017, 2019), for which the baseline length is comparable to that of the upcoming large sky surveys, such as LSST (see Table 1). It is important to point out that these data have specific characteristics which can be very difficult to simulate. For example, NGC 5548 shows the time-lag variability, Arp 102B and 3C 390.3 are classified as double peaked line (DPL) emitters but their oscillatory characteristics

Table 1. Summary of characteristics of the objects. The columns are: object name, monitoring time-baseline, AGN taxonomy, mean relative error in flux measurements (%), variability parameter (%), redshift, rest-frame time lag (corrected for time dilation), mean sampling time, error of rest-frame time lag, rest-frame oscillation period, error of the rest-frame oscillation period. As reference, H β line was used for spectral time lag measurement.

object	Base[yr]	type	σ [%]	F_{var} [%]	z	τ_{rest} [ld]	Δt [days]	err τ [%]	P_{rest} [yr]	err P_{rest} [%]
NGC 3516	22	CLAGN	4	15.8	0.0088	9.6	69.8	204.35	-	-
NGC 7469	19	Sy1.0	5	23	0.0163	20.7	32.2	32.8	7.006	41.46
E1821+643	24	quasar	5	7	0.297	90.98	64.3	0.065	9.84	33.84
Arp 102B	26	DPL	5	21	0.0242	14.65	96.9	130.2	-	-
Ark 564	11	NLSy1	5	7	0.0247	3.9	40.5	658.7	-	-
3C 390.3	12	DPL	5	38	0.0561	90.9	128.6	27.6	9.56	0.94
NGC 4151	10	Sy1.5-1.8	5	42	0.0033	4.98	25.3	558.2	13.71	27.02
NGC 5548	6	Sy1.0-1.8	5	33	0.0172	48.2	28.35	38.1	13.075	16.71

are quite different (Kovačević et al. 2018), E1821+643 shows extremely low variability as a binary SMBH candidate (Shapovalova et al. 2016), Ark 564 is as a narrow-line Sy 1 (NLSy1) object with very low-variability. In Shapovalova et al. series of papers (see e.g., Shapovalova et al. 2016), we introduced the time lag determination based on Gaussian process light curve modeling as a novel tool. Moreover, the periodicity detection for several of these objects was made possible by our new tool - 2DHybrid method (Kovačević et al. 2018). Since commonly used periodicity detection methods are not designed for red noise light-curves we applied our 2DHybrid method, which results are given in Table 1 (see also Kovačević et al. 2018). Some notable features of our 2D hybrid method are: enhancement of apparent resolution by spreading peaks over the second dimension, and establishment of direction of changes in signal through correlation coefficients (Kovačević et al. 2020). Even though some data was added from other RM campaigns, the Shapovalova et al. dataset served as the backbone for periodicity detection (Kovačević et al. 2018). Table 1. lists the object basic information and spectral characteristics calculated for the H β line: object name, monitoring time-baseline in years, AGN taxonomy, mean relative error in flux measurements σ (in %), variability parameter F_{var} (in %), redshift z , rest-frame time lag (corrected for time dilation) in light-days τ_{rest} , mean sampling time in days δt , relative error of rest-frame time lag in percentages err τ , rest-frame oscillation period in years P_{rest} and its relative error in percentages err P_{rest} . Two parameters, the level of variability and relative photometric error, used for the calculations of the statistical proxy given in Eq.(1) are taken from the Shapovalova et al. campaign. For the parameter F_{var} we use the variation amplitude with respect to the mean flux F_{mean} , defined as:

$$F_{\text{var}} = [\sqrt{\sigma(F)^2 - e^2}] / F_{\text{mean}}$$

e^2 being the mean square value of the individual measurement uncertainty $e(i)$ for N observations, i.e. $e^2 = \frac{1}{N} \sum_i e(i)^2$ (O'Brien et al. 1998). We note that other definition of the level of variability can be also used.

2.3.2. Artificial set of light curves with ideal and LSST OpSim cadences

Kelly et al. (2009) found that the optical variability could be represented by a stochastic model based on Damped random walk (DRW) process. The model incorporates a characteristic amplitude σ , which affects exponentially-decaying variability with time scale τ around the mean magnitude m_0 . The model specifications σ and τ are related to the SMBH mass M_{BH} and/or luminosity L of the AGN (Kelly et al. 2009, 2013). AGN prop-

erties are approximated as follows: luminosities are chosen randomly from the range $\log L \in [42.2, 47]$. The SMBH mass M_{BH} is determined by L , Eddington luminosity

$$L_{\text{Edd}} = 1.25 \cdot 10^{38} \frac{M_{\text{BH}}}{M_{\odot} [\text{erg s}^{-1}]}$$

(Woo & Urry 2002) and an Eddington ratio (Shankar et al. 2009). Then, the characteristic radius of the BLR is approximated by the empirical radius-luminosity relationship (Bentz et al. 2013). For the simulations, DRW scales (σ and τ) are drawn based on luminosity from the distributions given in Equations 22 and 25 by Kelly et al. (2009).

The sequence of AGN light curve points p_i comes from the DRW model which is recursive in the flux dimension and iterative in the time dimension (Kelly et al. 2009):

$$p_{i+1} = \mathcal{G}\left(p_i e^{\frac{-\Delta t}{\tau}} + m_0(1 - e^{\frac{-\Delta t}{\tau}}), \sigma \sqrt{\frac{\tau(1 - e^{\frac{-2\Delta t}{\tau}})}{2}}\right) \quad (6)$$

where \mathcal{G} is the Gaussian distribution, $\Delta t = t_{i+1} - t_i$ is the time interval, $m_0 = 23$ mag, $p_0 = \mathcal{G}(m_0, \sigma \sqrt{\frac{\tau}{2}})$ at t_0 .

For photometric uncertainty we adopt the photometric error model of LSST (Ivezić et al. 2019):

$$\sigma_{\text{LSST}}^2 = \sigma_{\text{sys}}^2 + \sigma_{\text{rand}}^2 \quad (7)$$

where $\sigma_{\text{sys}} = 0.005$ is the systematic error due to imperfect modeling of point source,

$$\sigma_{\text{rand}} = \left(\frac{1}{25} - \gamma\right)X + \gamma X^2$$

is the random photometric error where $X = 10^{0.4(m-m_5)}$, $m_5 = 24.7$ and $\gamma = 0.039$ for r -band (see Ivezić et al. 2019, and their Table 2). Finally, the observed light curve is obtained from

$$y_i = p_i + \mathcal{G}(0, \sigma_{\text{LSST}}(p_i)). \quad (8)$$

The parameters τ , σ and period of the simulated light curves are corrected for the $(1+z)$ where z is the redshift. The simulation of the flux in some emission line l , which is emitted by the BLR, is based on the linear approximation:

$$f^l(t) = (f^c * \xi)(t) = \int_{-\infty}^{\infty} f^c(\tau) \xi(t - \tau) d\tau \quad (9)$$

where the transfer function ξ defines, essentially, the geometry of the BLR region as seen by the observer and f^c is a flux

originating closer to the SMBH and driving f^l . For simplicity, we consider that $\xi \propto \mathcal{G}(R_{\text{BLR}}, 0.25R_{\text{BLR}})$ (Chelouche & Daniel 2012), where R_{BLR} is already derived from the mass of SMBH, and the input continuum flux f^c is the realization of Eq. 6. Unlike the spectroscopic approach, this method does not allow the emission line and continuum light curves to be extracted, since the data consists of their combined signal. In reality under certain condition, separation of these processes is possible so the lag can be measured (Chelouche & Daniel 2012). Particularly, we assume that continuum and emission line through given filter have been determined in advance either using Chelouche & Daniel (2012) method or known somehow else. Thus, we simulated disentangled continuum and emission line using Eq. 8 and Eq. 9, respectively, while taking into account dimension of BLR inferred at the beginning of our procedure.

For simulating underlying periodic signal, we assumed that the inferred SMBH mass is the total mass of the hypothetical binary system at mutual distance of ~ 10 ld and that the amplitude is about 14% resembling the case of PG1302-102 (D' Orazio et al. 2015). Also the modulation of the signal can be approximated to the first order by amplitude of $v \cos \psi \sin i/c$ where ψ is the orbital phase, v is the velocity of the secondary component and i is the inclination angle (D' Orazio et al. 2015). For near-equal mass binaries, some studies show that the mass accretion rates fluctuate periodically, but they resemble a series of sharp bursts, differing from sinusoid-like shape (D' Orazio et al. 2015).

These continuous artificial light curves were sampled according to several observing strategies:

- (i) Idealized observing strategies were constructed to assist in the interpretation of the results related to SF which are obtained from LSST-cadences, serving as approximations ranging from reasonable to excellent observing strategies:
 - (1) ideal light curves cadence: uniform (1-day cadence) during 10 yrs;
 - (2) "gappy" light curves cadence: series of 3 months /6 months/9 months of uniform (1-day cadence) observations per year during 10 yrs campaign. There are no observations in gaps;
 - (3) variable-cadence light curve: in the first year only three months are observed with 1-day cadence, and in the next years 3 months with 1-day cadence are observed, followed by 6 months of 30-day cadence, and a gap of 3 months.
- (ii) OpSim runs with three different observing strategies:
 - (I) observing some selected LSST fields by taking around 90 epochs during 10 yrs,
 - (II) observing fields with about 1500 epochs during 10 yrs,
 - (III) observing fields with 200 epochs over 10 yrs of survey.

Figure 1 (left panel) shows a realization of the artificial light curve with a periodic signal of 4.3 yr (see Eq. 8) and different cadences. The upper panel shows a light curve with uniform 1-day cadence case i-(1), and the bottom three panels give light curves with variable cadence case i-(2). The generated artificial light curves with variable cadence (denser and sparser, as described above) are shown in the left panel in Figure 2.

For further testing of the proposed metrics on the LSST observing strategies, we created artificial set of objects based on OpSim realizations (their designations listed in the last column of Table 2).

We selected several OpSim runs which are thought to be relevant for AGN research, such as the AGN Deep Drilling Fields (DDF), but in general, this concept could be applied on any OpSim run, therefore we have also used rolling cadences. Rolling

cadences are non-uniform observing strategy where some region of the sky is emphasized in one year, and then minimized in the next.

The DDF simulations have different observing cadence for the five DDFs, while a standard baseline observing strategy is applied for the rest of the sky. The AGN DDF OpSim run takes shorter DDF sequences more often (around 2.5 % of visits are spent on DDFs)¹.

We used FBS 1.5² AGN DDF runs in g and r bands for testing the properties of structure function. The r band has a total number of 491676 visits during the 10-year period of LSST mission, while the g has 223871 visits.

In Figure 3 the total number of visits per equatorial coordinate is shown for the r band. The noticeable difference between number of the visits in DDFs and the rest of the sky is clearly visible. This allow us to choose one point (for definition of point see, e.g. Biswas et al. 2020) from DDF and one point outside the DDF in order to probe the difference.

We demonstrate our analysis using two different OpSim cadence releases, OpSim 1.5 and 1.6. From the OpSim 1.5 we used AGN DDF cadences and from the OpSim 1.6 we used rolling cadences. Results based on the newest cadence release, OpSim 1.7, can be found at our online supplementary channel (https://github.com/LSST-sersag/agn_cadences). Also, our repository could be efficiently used for different OpSim realisations. There are no discrepancies between results obtained from these two realisations.

Based on described procedure, we generated the set of artificial objects using the cadences from the OpSim. Further, we applied our methods to determine the time lag and periodicity. Both, the model input and measured values are given in Table 2. Using these values we performed multiple linear regression as in the case of our true monitoring campaign. Table 2 lists the variability characteristics of the artificial light curves, i.e.: object ID, luminosity, SMBH mass, BLR dimension, mean error in magnitude measurements (in %), variability parameter (in %), redshift, imparted period of oscillations in the light curve given in rest frame assuming that the mass is a total mass of binary at mutual distance of 10 ld, cadence from the OpSim realization, detected rest-frame time lag, error of rest-frame time lag, rest-frame oscillation period, error of the rest-frame oscillation period, designation of the OpSim realization.

3. Results

The results are presented in three main sections that seek to capture the different perspectives of the relation between cadence estimates and AGN variability-related observables. A more general issue of the number of binary AGN candidates detected by LSST like survey allows us to infer general constraints on cadences compatible with multiple regression predictions. The cadence estimates for time-lag and oscillation measurement use real -world and LSST-like objects samples. To understand the factors affecting cadence estimates for SF, we tested SF-metric with idealized and LSST-like data set.

3.1. Cadence estimates for time-lag measurement

A success of the recovery of time lag and periodicity depends on the observing cadence (Horne et al. 2004) as well as the cadence of observations. We thus learned the multiple regression of the

¹ <https://pstn-051.lsst.io/PSTN-051.pdf>

² <http://astro-lsst-01.astro.washington.edu:8081>

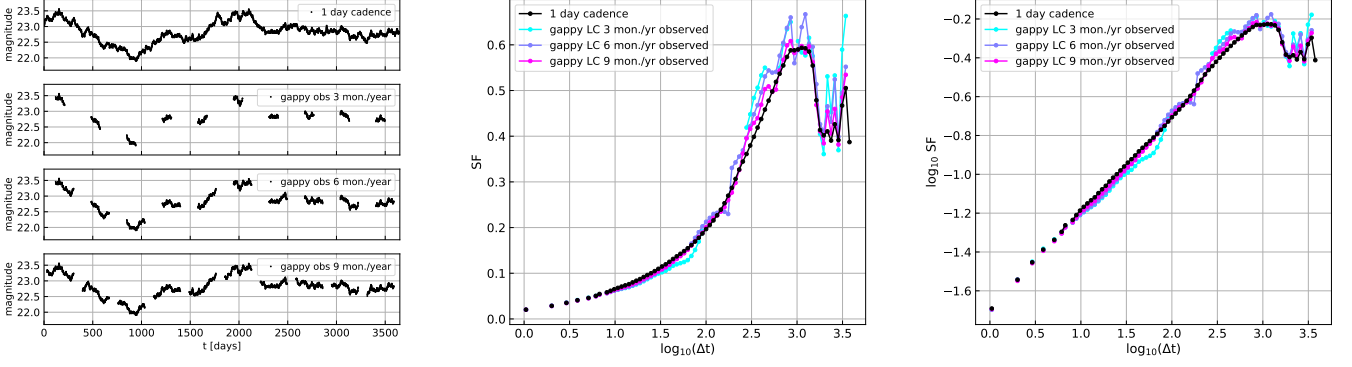


Fig. 1. *Left:* Artificial AGN light curve based on DRW model with added oscillatory signal. Generated light curve corresponds to AGN with black hole mass of $2.8 \times 10^8 M_\odot$ and bolometric luminosity of $8.5 \times 10^{45} \text{ erg s}^{-1}$. The period is 4.3 yr. Upper panel gives the ideal light curve with homogeneous 1-day cadence, the bottom three panels give "gappy" light curve assuming respectively that 3, 6 or 9 months per year are observed. For all idealized rolling cadences during observed periods the sampling rate is 1 day. *Middle:* SFs calculated for the light curves given in left panels. Black curve is the SF calculated for the ideal light curve with homogeneous 1-day cadence. Blue, violet, and pink stand for SFs calculated for gappy cadences of 3 months/yr, 6 months/yr and 9 months/yr respectively. *Right:* The same plot as in the middle panel but in logarithmic scale.

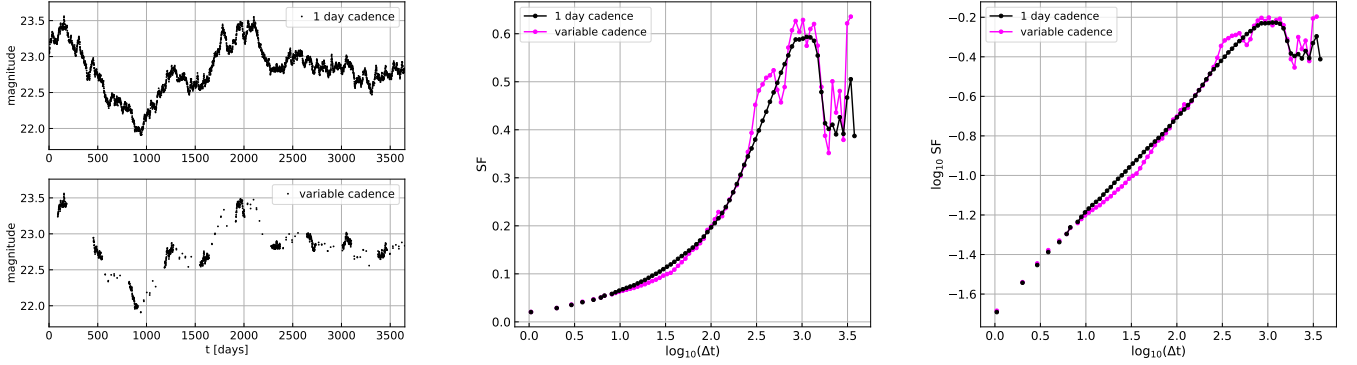


Fig. 2. *Left:* Artificial AGN light curve based on DRW model with added oscillatory signal, but now with variable cadence. Prior to light curve generation itself, we used the same seed in the pseudorandom number generator to obtain the exact same light curve as the one in Figure 1. Upper panel gives the ideal light curve with uniform 1-day cadence, and the bottom panel gives the variable-cadence light curve assuming that 3 months per year are observed with 1 day cadence and 6 months per year with 30-day cadence. *Middle:* SFs calculated for the light curves given in the left panel. Pink curve is the SF of variable-cadence light curve, and black stand for the SF calculated for ideal light curve with uniform 1-day cadence. *Right:* The same plot as in the middle panel, but in logarithmic scale.

Table 2. Summary of characteristics of the artificial objects. The columns are: object ID, luminosity, SMBH mass, BLR dimension, mean relative error in magnitude measurements (in %), variability parameter (in %), redshift, imparted period of oscillations in the light curve given in rest frame assuming that the mass is a total mass of binary at mutual distance of 10 ld, cadence, detected rest-frame time lag, error of rest-frame time lag, rest-frame oscillation period, error of the rest-frame oscillation period, designation of the Opsim realization.

ID	L [$10^{44} \text{ erg s}^{-1}$]	M_{BH} [$10^6 M_\odot$]	R_{BLR} [ld]	σ [%]	F_{var} [%]	z	P_{rest} [yr]	Δt [days]	$\bar{\tau}$ [ld]	$\delta\bar{\tau}$ [ld]	\bar{P}_{rest} [yr]	$\delta\bar{P}_{\text{rest}}$ [yr]	OpSim_cad
I	85.5	282.7	36.03	0.16	14.9	0.05	4.28	43	39.05	5	3.93	1.4	roll_cad_0.8_g_ra_0_de-10
II	1156.7	1605.4	144.5	0.23	09.3	0.05	1.8	1.4	138.1	4.8	1.85	0.31	agn_g_ra_9.0_de-44 (DDF)
III	199	496.5	56.5	0.19	16.2	0.05	3.23	18	54.3	5	3.2	0.45	roll_cad_0.8_ra_0.0_de-30
IV	80.4	271.3	34.87	0.14	20.4	0.05	4.4	17	31.4	4.7	4.4	0.5	roll_cadence_0.8_ra_0.0_de-50
V	12.1	76.6	12.68	0.22	18.8	0.05	8.2	18	16.1	4.9	6.9	0.7	roll_cad_0.8_r_ra_0.0_de-30
VI	391	778.8	81.02	0.23	15.3	0.05	2.58	19	84.5	4.8	2.12	0.31	roll_cad_0.8_z_ra_0.0_de-10
VII	15	88.5	14.24	0.18	13.4	0.05	7.66	60	16.2	4.8	7.9	0.76	roll_cad_0.8_u_ra_0.0_de-30

proxy variable for the relative uncertainty in time lags given by Eq. 1 and plot its relationship with two independent variables (see Figure 4 left plot) based on our data given in Table 1. The multiple linear regression (blue plane) is tilted along both independent variables. In other words, the proxy of uncertainty is decreasing along the first and increasing along the second inde-

pendent variable, corresponding to the first and second term on the right hand side of Eq. 1. For this analysis, the time lags have been restricted to the lags of $H\beta$ only. Each object has the same weight in the regression but we do not, in fact, expect the luminosity and lag to be the same in campaigns during different years as for NGC 5548. The left panel of Figure 4 shows each individ-

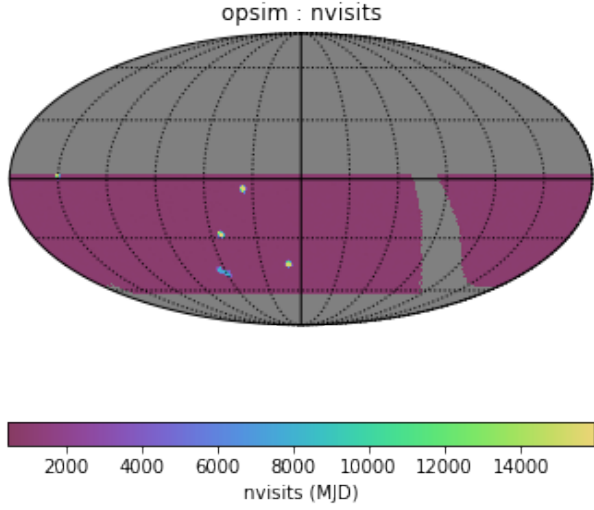


Fig. 3. Number of visits for each equatorial coordinate pair during the whole scope of the LSST mission (10-year period) for the r band. Color bar represents number of visits. DDF fields are easily distinguished with largest number of visits (yellow and blue color). Plot is obtained using the Python-based LSST Simulation Framework.

Table 3. Comparison of the prediction of cadences from models derived from empirical Δt_E and artificial Δt_A data sets, assuming the flux errors of 5%, 0.01%, respectively; and time lag uncertainty $\sim 10\%$. Luminosities $L[\text{erg s}^{-1}]$ are calculated from given fiducial time lags τ using the R-L relation (Bentz et al. 2009b).

$F_{\text{var}}[\%]$	$\tau[\text{ld}]$	$L[\text{erg s}^{-1}]$	z	$\Delta t_E[\text{days}]$	$\Delta t_A[\text{days}]$
10	100	$5.8 \cdot 10^{46}$	0	63	83.9
			1	31.2	48.9
			4	12.6	20.1
			7	7.9	12.4
	10	$7.75 \cdot 10^{44}$	0	6.2	9.9
			1	3.2	4.83
			4	1.3	1.89
			7	0.8	0.9
20	100	$5.8 \cdot 10^{46}$	0	95	110
			1	47.7	95
			4	19.1	37.2
			7	11.9	23.5
	10	$7.75 \cdot 10^{44}$	0	10	12.9
			1	4.8	9.64
			4	1.9	3.8
			7	1.2	2.2

ual data point for the relative uncertainty in time lag from the monitoring campaigns included here. Blue plane shows the best multiple-linear regression model to the relationship given by Eq. 1.

One can now ask the question: What would be a suitable cadence for detecting certain AGN observable at the level of formal error of 10% if setting the fiducial light curve variability and flux error. Based on the multiple regression model best-fit of empirical data (E) and artificial data (A), and for the time lag error of 10%, we can vary parameters in both independent variables in order to get cadence estimates.

The resulting prediction (see Table 3) shows detailed information on each cadence which includes assumed variability, redshifts, time lags and luminosities. A close inspection of Table 3

Table 4. The same as Table 3 but for the fiducial luminosities $L[\text{ergs s}^{-1}]$ at given redshift (Kollmeier et al. 2006) while fiducial time lags τ are obtained using the R-L relation (Bentz et al. 2009b).

$F_{\text{var}}[\%]$	$\tau[\text{ld}]$	$\log L[\text{ergs s}^{-1}]$	z	$\Delta t_E[\text{days}]$	$\Delta t_A[\text{days}]$
10	11.5	45	1	3.67	5.8
	39.2	46	2	8.33	13.36
	72.4	46.5	3	11.45	18.29
20	11.5	45	1	5.6	10.9
	39.2	46	2	12.7	24.4
	72.4	46.5	3	17.5	34.2

Table 5. The same as Table 3 but for the periodicity.

$F_{\text{var}}[\%]$	$P[\text{yr}]$	z	$\Delta t_E[\text{days}]$	$\Delta t_A[\text{days}]$
10	5	0	29.2	80.3
		1	14.6	62.1
		4	7.3	47.5
		7	3.7	32.8
		0	18.3	62.1
	3	1	11	54.7
		4	4	36.5
		7	2.2	25.6
	5	0	91.3	92
		1	62.1	73
20	4	0	25.6	29.2
		7	14.6	18.25
		0	65.7	69.4
	3	1	36.5	40.2
		4	14.6	18.25
		7	9.1	11

shows that required cadences are smaller for objects with larger redshift, and for smaller time lags as expected. Also, two times larger object light curve variability ($F_{\text{var}} \sim 20\%$) allows larger cadences for two different fiducial time lags (luminosities) of potential targets.

Cadence of the order of tens of days is qualitatively sufficient for time lag estimates of 100 days for light curves of smaller and larger variability. Also, ten times smaller time lags would require ten times smaller cadence. A relationship between AGN luminosity and redshift (L - z) can also be combined with the R-L relation to give a rough estimate of the required cadence for different luminosities up to $z \sim 3$ (Table 4). For given redshift, the luminosity is taken from empirical L - z relation reported in (Kollmeier et al. 2006), while R-L relation (Bentz et al. 2009b) provides corresponding time lag. It is expected that with larger lags (i.e. luminosities) cadences increase. Similarly, we estimated the proxies (see the left panel in Figure 5) for the set of data (A) based on OpSim cadences (see Table 2, last column). Cadence predictions obtained from the two model versions are of the same of order so that the two models are in relative good agreement (see Table 3 and 4, last columns).

Even though LSST OpSim cadences have gaps, we recovered lags which are consistent within 3σ with the R_{BLR} of the input model. The success rates can be boosted by using deep learning to deal with gapped light curves (Faisst et al. 2019).

To compare multiple regression model prediction of $\log \frac{\sigma_T}{T}$ with those values obtained from data sets we used density estimator. Such estimator is an algorithm which takes a dataset and produces an estimate of probability distribution which that data is drawn from. The inferred distributions from model results and data we will call descriptors and they are compared

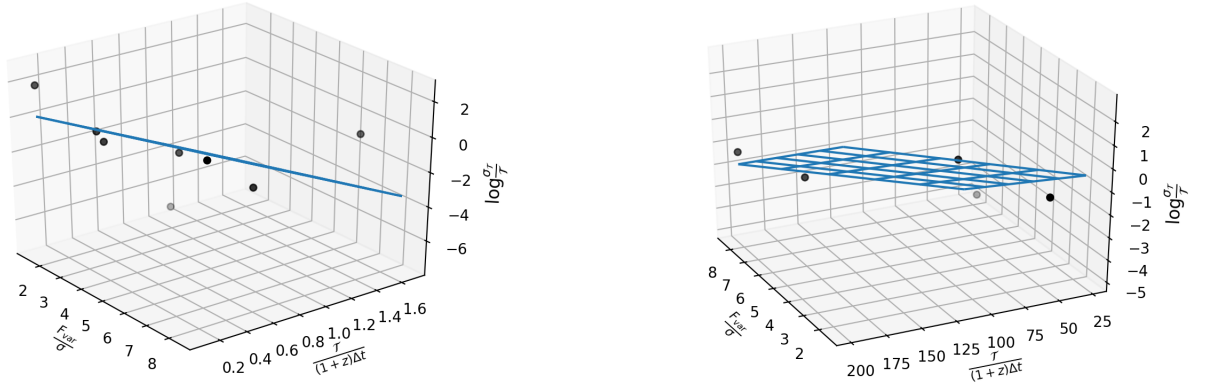


Fig. 4. 3D plot of fitted multi-linear regression model (blue plane) for empirical proxies for uncertainties in time lags (left) and in periodicities (right). Filled circles denotes data given in Table 1. *Left* $C_1 = 0.34 \pm 0.20$, $C_2 = -3.75 \pm 1.38$; *Right* $C_1 = -0.64 \pm 0.13$, $C_2 = 0.018 \pm 0.0068$.

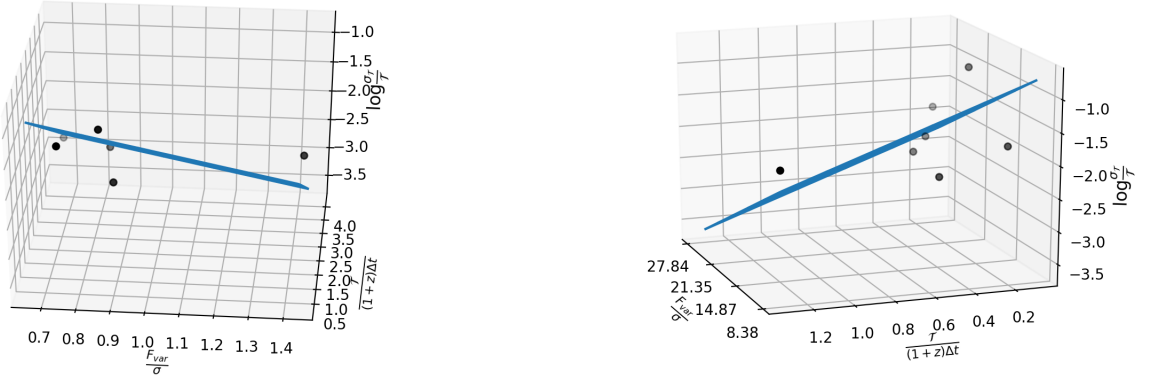


Fig. 5. Same as the comment for Figure 4 but for artificial set of light curves with OpSim cadences. Filled circles denote data given in Table 2. *Left*: $C_1 = -1.09 \pm 0.29$, $C_2 = -0.49 \pm 0.12$, σ is of the order of 0.01; *Right*: $C_1 = -0.0008 \pm 0.0001$, $C_2 = -1.089 \pm 0.5$, σ is of the order of 0.005.

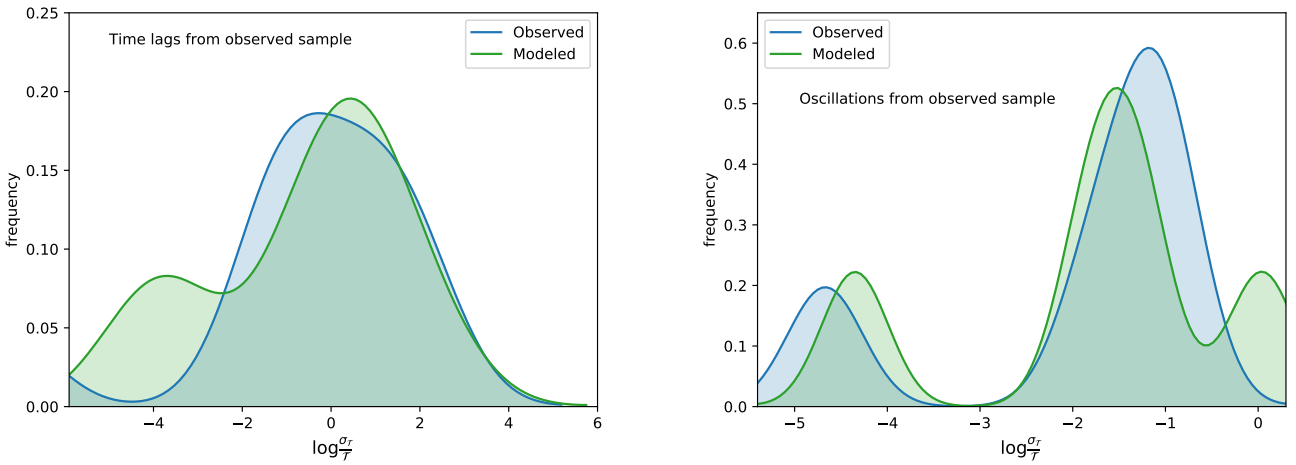


Fig. 6. Probability density functions of formal errors from observed data and their predicted values from multiple regression for time lags (left) and oscillations (right).

against each other. Particularly, we implemented Kernel density estimation (KDE) which uses mixture consisting of one kernel component per point in the considered data set, resulting in an essentially non-parametric estimator of density (see e.g., Chen 2017). There are several versions of kernel density estimation implemented in Python (notably in the SciPy and StatsModels packages). The density estimate at a point y within a group of points $\{x_i\}$, $i = 1, N$ is given by:

$$f_K y = \sum_{i=1}^N K(y - x_i; h) \quad (10)$$

where in our case

$$K(x; h) \propto e^{-\frac{x^2}{2h^2}}.$$

The left plot in Figure 6 displays the probability density of formal errors of time lags inferred from the observed data and from multiple regression predictions. Larger discrepancies can occur in the left tail of observed formal errors. Similarly, the left plot in Figure 7 shows the same information but for artificial data set. The model much better perform on this data set.

3.2. Cadence estimates for oscillation detection

Now we consider the cadence required for reliable detection of oscillation, i.e. periodicity in light curves. Here we will repeat the procedure as for the time lag cadence. The relationship between the uncertainties for detected periods (see Table 1) and proxy variable (Eq.1) is shown on the right panel of Figure 4.

Full details of the predicted cadences for two hypothetical oscillations (3 and 5 years) in decade long light curve are given in Table 5.

For larger number of cycles of underlying signal, the predicted cadences are somewhat smaller. For example, the light curve with $F_{var} \sim 20\%$ and underlying rest-frame periodicity of 3 years at redshift $z \sim 4$ would require cadence of ~ 14.6 days. Such cadence would be sufficient for detection of a 5 year rest-frame oscillation (at similar level of light curve variability $F_{var} \sim 20\%$), for an object at $z = 7$ (see Table 5).

We performed multiple regression (see right panel in Figure 5) on the set of artificial objects designed to have cadences from several OpSim runs (see Table 2).

While results of models based on artificial and empirical data are in good agreement for $F_{var} \sim 20\%$, there are considerable differences for $F_{var} \sim 10\%$, (details in Discussion section).

The right plot in Figures 6 shows the probability density of formal errors of oscillations inferred from the observed data and from multiple regressions. The distribution is multimodal, but model is relatively close to the measured formal errors. However, the distribution of artificial errors is bimodal (not multimodal as in the case of observed data). The distribution of predicted formal errors is broadened to capture both peaks of measured errors. The right plot in Figure 7 shows the same but for artificial data set.

3.3. Structure Function Results

SF, as a classical method, has been used to detect the periodicity and timescale of AGN variability in different observing bands (Wang et al. 2017; Moreno et al. 2019). It is believed that SF is suitable to handle unevenly sampled time series data. Thus, we also analyze the influence of different cadences on the first-order SF method applied on AGN light curves with underlying

oscillations. The ideal and LSST-like cadences were used for SF construction.

3.3.1. Simulated cadences

Firstly, we investigate individual SFs for the "gappy" artificial light curves (left panel, Figure 1) and variable cadence (left panel, Figure 2), in order to compare cadence effects.

Right panel in Figure 1 shows that the SFs of the "gappy" light curve with continuous observations of 9 months (red), 6 months (purple), 3 months (cyan), with respect to the SF of the reference light curve with homogeneous 1-day cadence (black). The SF for 3 and 6 months of observations shows a large deviation SF of the ideal light curve of homogeneous 1-day cadence.

For variable-cadence light curves, we can expect larger deviations of SF than in previous case (see middle panel in Figure 2), but periodic signal is clearly visible at the largest timescales of the SF in comparison to the ideal light curve with homogeneous 1-day cadence. The right panel shows the same SF as in the middle panel but on logarithmic scale. Deviations at smallest timescales appear, whereas oscillatory patterns are persistent, but less prominent.

Further, we consider deviations of SFs of an ensemble of homogeneous light curves across redshifts. For each redshift bin ($z_{bin} = \{i | i \in [0.5, 6.5], \Delta i = 0.5\}$), we generate a 10-year long artificial light curve based on DRW model, setting continuous 1-day cadence and underlying oscillatory signal according to described procedure. Then, we apply to these light curves the "gappy" observing strategy with 3 months/yr, 6 months/yr and 9 months/yr of observations of different cadence: 1, 4 and 10 days. This would generated 50 light curves for each redshift bin and three "gappy" observing strategies. For each "gappy" light curve, we estimate the SF_{gappy} . For each redshift bin, we derive the averaged deviation curve M^i . Then we plot M^i vs. redshifts z with projected SF timescales Δt in the form of heatmaps. Below we discuss the results of these three observing strategies.

Case i) The heatmaps of SF deviations for observing strategy of 3 months/yr are given in Figure 8 for sampling rates of 1 day (left panel), 4 days (middle panel) and 10 days (right panel). The deviations form 'evolutionary tracks' for time scales between 1 and 3.0 are seen. At smaller scale (below 1.5), SF deviations became apparent while it becomes 'noisy' when the sampling rates are larger (4 and 10 days - middle and right panel, respectively). Changes at larger time-scales ($\Delta t > 2$) are prominent also for larger cadences (4 and 10 days) at larger redshifts (middle and right panel in Figure 8). We note that deviations can be either positive or negative following mild bent tracks across redshifts. Positive deviations mean that SF values of continuous curve are larger than SF of "gappy" curve and vice versa. Deviations became more 'noisy' at intermediate scales (1,2.5) for sparse sampling of 10 days. Dominant positive deviation (values in the range (0,2)) seen as blue track is persistent across redshifts and different sampling rates.

Case ii) For comparison, heatmaps of SF deviations for "gappy" strategy of continuous observations during 6 month/yr are given in Figure 9. The maps are smoother then corresponding maps for case i) cadence (Figure 8). Also, the evolutionary tracks of deviations are translated to larger time scales beyond 2.5. The deviations at smaller scales (below 1.5) became apparent and become noisy with larger cadence of 4 and 10 days. Deviations are smaller then in the case i). The blue track dominant for case i) "gappy" observing strategy, disappears from heatmaps of 6month/yr "gappy" observing strategy.

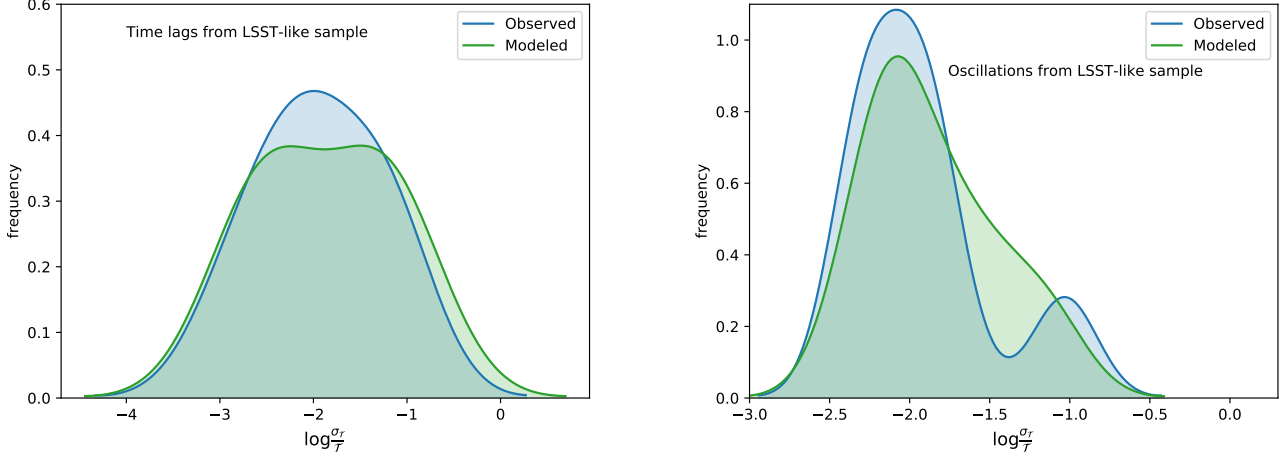


Fig. 7. The same as Figure 6 but for artificial data which are also referred as 'observed'.

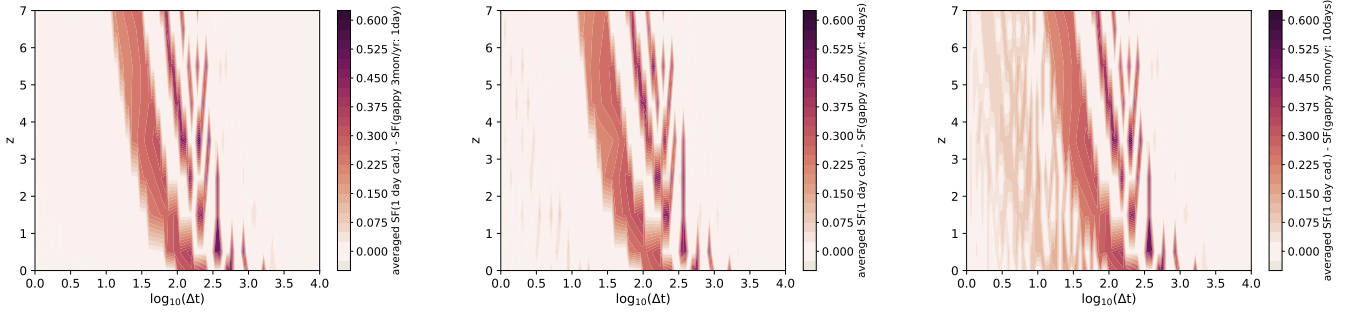


Fig. 8. Heatmaps of deviation of SFs for cadence 3month/yr in the rest frame of quasar. From left to right: the sampling rate is 1 day in observed 3months/yr; 4 days sampling rate in observed 3months/yr; and 10 days sampling rate. Colorbar represents deviations. Positive deviations stand for SFs when values of homogeneous curve are larger than SF of gapped curve in average per bin and vice versa.

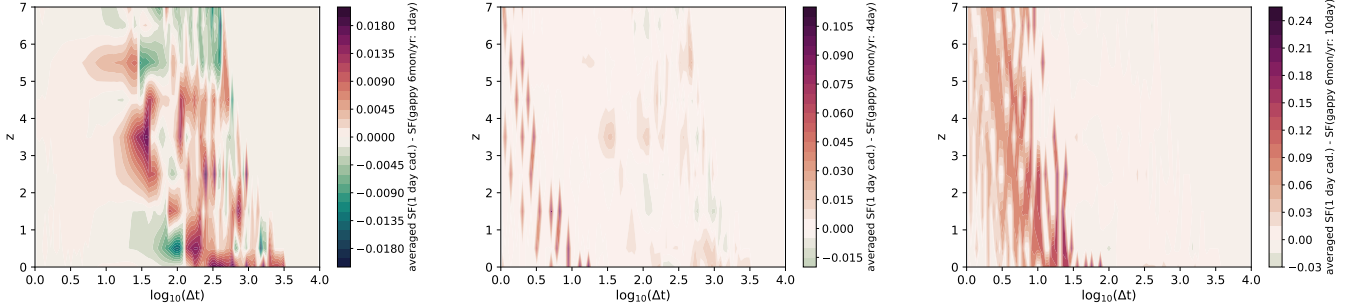


Fig. 9. The same as Figure 8 but for cadence 6month/yr in the rest frame of quasar.

Case iii) For 9 month continuous observations, heatmaps in Figure 10 are similar to the corresponding heatmaps for case ii), but deviations are smaller. The evolutionary tracks of differences are attenuated at larger time scales. As expected, the noise appears at smaller time scales for larger samplings.

For variable-cadence light curve comprising 3 observed months with 1-day sampling and 6 months with 30-day sampling, with exception of the first year when the ideal 1-day cadence during 3 months is present, the heatmap is given in Figure 11. The evolutionary track of deviations at time scale around 2.5 is prominent. The noise is present at intermediate time scales between 1 and 2. We note that it is similar to the heatmap of 6-

month "gappy" observing strategy with 1-day sampling but with reversed coloring (the left plot in Figure 10.) It means that denser sampling within 3 months helps to get smaller deviations of SF of gapped light curves.

3.3.2. OpSim Rolling cadences

We generated artificial light curves with underlying oscillations that correspond to different cadences runs from OpSim outputs. OpSim_roll.cad_0.8_g_RA_0.0_D_-10.0 comprises 87 and OpSim_roll.cad_0.8_r_RA_0.0_D_-10.0 204 observations over 10 yr. A realization of these light curves is given in left panel of Fig-

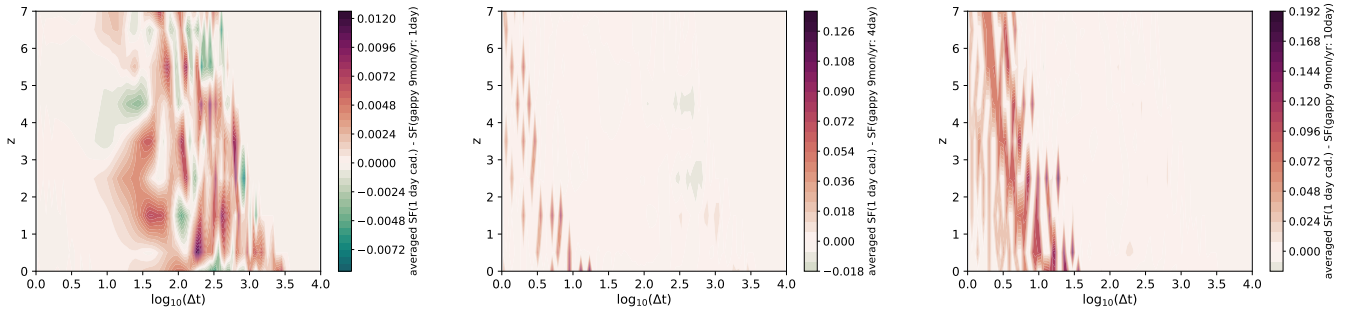


Fig. 10. The same as Figure 8 but for cadence 9month/yr in the rest frame of quasar.

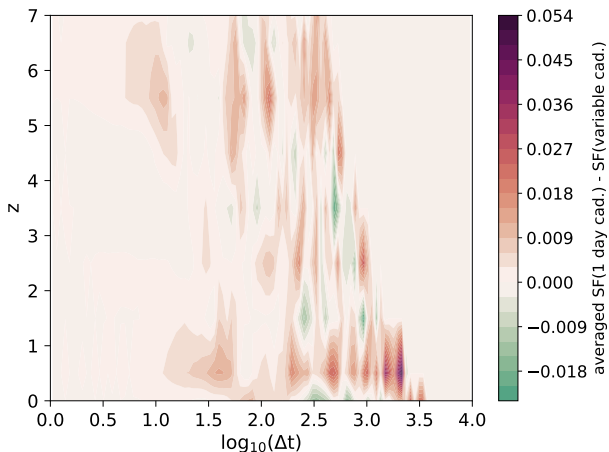


Fig. 11. The same as Figure 8 but for variable cadence comprising 3months observed with 1 day sampling followed by 6 months with 30 day sampling, with exception of the first year when the cadence of 3 month is present.

ure 12, for g and r filter OpSim rolling cadence. Corresponding SFs (right panel in Figure 12) show larger deviations of SFs in g and r filter at time scales below 1.5, oscillations are present after time scale 2.0. In g filter, a dip of the SF is around time scale 2.25. SF in r filter follows closely SF₀.

Averaged SF deviations for OpSim1.5 deep drilling field (DDF)_AGN.g_ra_0.0_de_-30.0 and AGN.r_ra_0.0_de_-30.0 are given in Figure 13. AGN.g_ra_0.0_de_-30.0 realization is fragmented and noisy, indicating erratic behavior of SFs. The evolutionary track close to the time scale 2.5 is present. AGN.r_ra_0.0_de_-30.0 produces heatmap which is similar to the simulated rolling cadence of 3 months/yr with 1-day sampling (Figure 8). The deviations evolutionary tracks are less fragmented and more bent, with two times larger values then those found in heatmaps of 3 months/yr gappy cadence. However, OpSim agn_g_ra_9.0_de_-44.0 cadence contains 2616 observations and produces map (right panel) similar to map 6 months/yr cadence Figure 9.

4. Discussion

To quantify LSST-like observing strategies' products, we focused on AGN variability-related observables (time lag, periodic oscillations, and SF) and their connection to predicting the most suitable LSST-like cadences. The importance of the first two lies

in their effects on constraining reliable AGN models and the third is essential for direct measurement of the covariance function of the AGN light curves and can display oscillatory signals. In this light, we propose a multiple regression model to statistically identify the cadence-formal error pattern knowing AGN-variability observables from surveys (real-world and simulated) operations.

In order to evaluate the performance of the proposed regression model, case studies of a real-world (pre-LSST era) and artificial observing strategies (LSST-like) are provided. Assumptions made in multiple regression model abstract from the details of the real and artificial light curves, yet capture the general relationship with AGN-variability observables. It connects the relative error of the AGN observable with light curve variability, flux errors, curve cadence and measured observable. It is reasonable to expect that the error of time lag and period will decrease with increasing $\frac{F_{var}}{\sigma}$ and $\frac{T_{obs}}{\Delta t}$.

Figure 4 shows the comparative multiple regression model prediction performances between empirical and artificial LSST-like data sets. The appearance of outliers is due to unfavorable combinations of formal errors, cadences, and gaps in the light curves. The time lags, periodicities, and their uncertainties are determined using Gaussian Process learned light curves from real and simulated ones. A simple inspection of time lags and periodicities obtained from real (Table 1) and artificial data (Table 2), show that the formal errors in real data are more fluctuating than those found in LSST like set. Possibly this is a consequence of real survey cadences having more random gaps than planned LSST-like strategies. The formal errors of detected periodicities are calculated as half-width of the relevant correlation cluster in 2DHybrid method, which depends on the amplitude of correlation peaks of continuous wavelet transforms of light curves (Kovačević et al. 2018). Also, some lower-luminosity objects such as NGC 4151 were, in general, targeted by the early ground-based monitoring campaigns due to their low redshift and apparent high brightness. The more considerable uncertainties in their time lag measurements partially come from the loosely constrained observational factors (such as observing cadence, spectral resolution, detector efficiency, etc.). Two versions of multiple regression model predict comparable time lags and oscillation cadences, while at $F_{var} \sim 10\%$ the results differ for oscillation detection. We propose that this can be explained by the sensitivity of oscillation detection to the AGN variability and light curves characteristics. The artificial set of light curves supports our general expectation from multiple regression model that the error of time lag and period will decrease with increasing two variables (coefficients C_1 and C_2 are negative). The empirical set of light curve produces model realization with alternate sign of

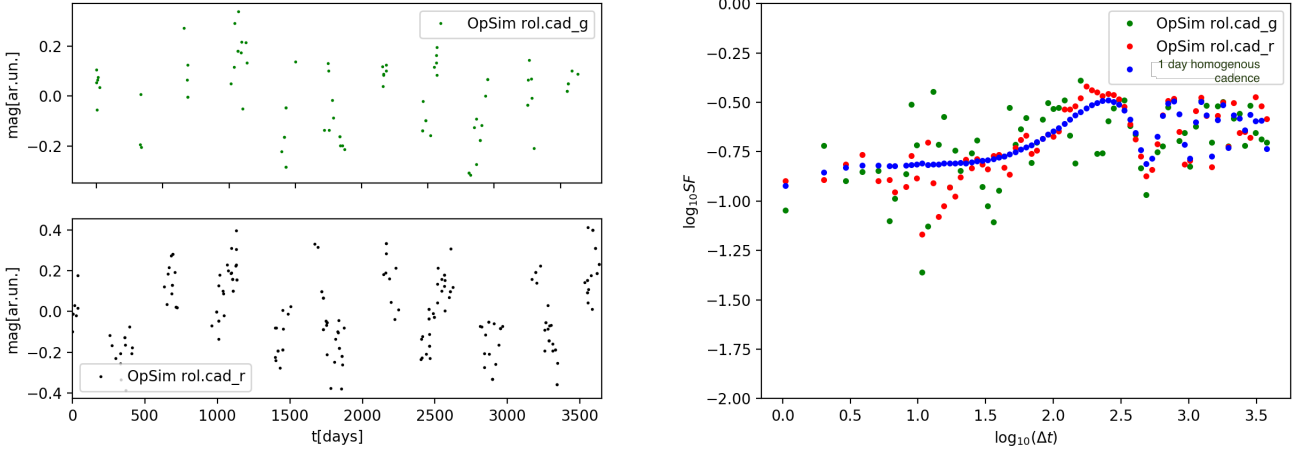


Fig. 12. *Left:* Simulated AGN light curve using DRW and OpSim rolling cadences. From top to the bottom: light curve with OpSim_rol.cad_0.8_g_RA_0.0_D_-10.0 cadence; light curve with OpSim_rol.cad_0.8_r_RA_0.0_D_-10.0. *Right:* Structure function calculated for the light curves given in the left panel. Blue curve stands for SFs calculated for homogeneous 1 day cadence.

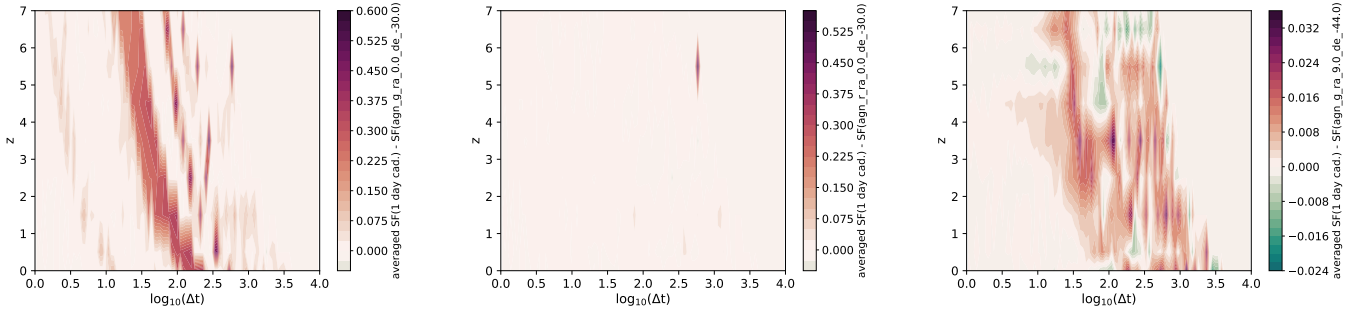


Fig. 13. Heatmap of SF-deviations for OpSim1.5 deep drilling field_ AGN.g_ra_0.0_de_-30.0 (left panel), OpSim1.5 deep drilling field_m_ AGN.r_ra_0.0_de_-30.0 (middle panel) and OpSim agn.g_ra_9.0_de_-44.0. Colorbar represents deviations. Positive deviations stand for $SF_0 > SF_{gapped}$ in average per redshift bin and vice versa.

coefficients C_1 and C_2 , which implies that there can be an additional systematic variable (or error) in the real data set.

While these findings are highly promising, our next steps will be testing the proposed regression model in much larger sets of objects such as from the SDSS RM campaign, which has monitoring 849 spectroscopically-confirmed quasars during 3-years long-period (Shen et al. 2019). Moreover, the concept shown here could be potentially important for designing the cadence strategy for the MSE quasars reverberation mapping survey of ~ 5000 quasars.

Some phenomena could affect positively or negatively the detection of the underlying oscillatory signal. For example, we analyzed periodicities for PG 1302-102 (Kovačević et al. 2019) and Mrk 231 (Kovačević et al. 2020) which were observed photometrically by Catalina Real-time Transient Survey (CRTS) and All-Sky Automated Survey for Supernovae (ASAS-SN). Analyzing Mrk 231 photometric curve, we found that adequate data sampling of ASAS-SN survey is more suitable for periodicity detection (Kovačević et al. 2020). Sudden changes in the target light curves can occur, as was the case of PG 1302-102 when unexpected flare appeared recently. This object has been considered as one of the best targets for the next generation of gravitational wave surveys. This demonstrates the importance of the information contained in individual segments of light curve. We emphasize that input for the artificial light curve based on OpSim

cadences differs from the real RM monitoring by not including factors such as flares, real physical processes, observation uncertainties, jets, etc. Since the magnitude of the regression coefficients depends on the scaling of the parameters are related not merely light curves parameters but also to mean cadence and the formal errors of derived quantities from the data sets, the coefficients of model runs on real and LSST-like surveys are not directly comparable and represent different model runs.

To further expand on the topic of possible periodicity detection, one can ask the question how many close binary SMBH systems could be detected by LSST.

Assuming that detection of periodicity in the light curves is possible if binary mutual separation is above of anticipated value $10\mu\text{as}$ (e.g., corresponding to the binaries at mutual separation ~ 0.01 pc and distance of $\sim 200\text{Mpc}$), and the orbital period is shorter than twice of the survey lifetime. The minimum binary separation a and the binary mass M give the minimum binary SMBH orbital period for which LSST could detect orbital motion:

$$P_{\min} = \frac{2\pi a_{\min}^{3/2}}{\sqrt{GM}}, \quad (11)$$

where is assumed that

$$a_{\min} = \theta/d \geq 10\mu\text{as}, \quad (12)$$

and P_{\min} and M depends on luminosity and redshift of targets, assuming that at angular-diameter distance d , the orbital angular radius of binary SMBH is $\theta \sim a/d$ where a is semimajor axis of binary.

We calculate the number of binary SMBH which can be detected by analyzing LSST light curves up to $z \sim 5$, using the estimated number of binary SMBHs per $\log z$ (D' Orazio & Loeb 2019):

$$\frac{dN}{d \log z} = 4\pi \frac{d^2 V}{dz d\Omega} \frac{\phi_0}{(\frac{L}{L_0})^{\gamma_1} + ((\frac{L}{L_0})^{\gamma_2} \min\{\frac{t_{res}}{t_l}, 1\}) (1 + e^{-2W})}, \quad (13)$$

where

$$\frac{d^2 V}{dz d\Omega}, \quad (14)$$

is the co-moving volume per redshift and solid angle (Ω). Also,

$$\frac{\phi_0}{(\frac{L}{L_0})^{\gamma_1} + (\frac{L}{L_0})^{\gamma_2}} \quad (15)$$

is the quasar luminosity function (see Hopkins, Richards & Hernquist 2007, parameters are given in the last row of their Table 3), where

$$t_{res} = \frac{20}{256} \left(\frac{P}{2\pi}\right)^{8/3} \left(\frac{GM}{c^3}\right)^{-5/3} q_s^{-1} \quad (16)$$

is the residence time of binary due to gravitational wave emission, $t_l \sim 10^7$ yr is the approximate AGN lifetime, $W = 10\text{yr} - P_{\min}$ where 10 yr is a LSST mission lifetime. For simplicity, we assume that at larger redshifts we expect brighter and more massive sources.

Figure 14 displays the distribution of detectable CB-SMBH of total mass $10^8 M_\odot$ and for three mass ratios $q = 1, 0.5, 0.05$, at different redshifts and cadences. The peak for all distributions is about redshift 2 as expected for AGN, however the number of possible detections varies across CB-SMBH mass ratios. It seems that as mass ratio decreases the number of possible detections increases. For example for mass ratio $q = 1$ and $q = 2$ we expect probability density function (PDF) peak at 20-25 objects (see Figure 14), however for $q = 0.05$ we can expect even 6 times more possible detections.

Although the number of expected CB-SMBH seems to increase with decreasing mass ratio, this does not necessarily mean that the number of effective CB-SMBH's detection increases. Namely, the decreasing mass ratio implies that the light curve will contain a periodicity signal with a smaller amplitude, but the small-amplitude oscillation detection is harder to perform. In reality, perhaps many factors can affect detectability. Some MHD studies (e.g., D' Orazio et al. 2013; Farris et al. 2014; Shi & Krolik 2015) have simulated unequal-mass ≤ 0.1 binaries. Their accretion rates are less bursty; and the cases for $q = 0.075$ and $q = 0.1$ binaries in D' Orazio et al. (2013) are very similar to PG 1302-102's light-curve, which has smooth sinusoidal appearance. Also, the cadences between 20 and 80 days (Figure 14) are sufficient for the most probable detections, which is in agreement with results given in Table 5.

The periodicity signal is present in SFs based on ideal and LSST-like cadences. The signal is prominent in the SFs for homogeneous 1-day cadence and in 'ideal surveys' "gappy" light curves with 6 and 9-months observing sets. The separation between subsequent wiggling of SF peaks is about imparted oscillation in the light curves. The deviation of SFs of light curves

with larger gap from the SF of the ideal series with homogeneous cadence inspired us to introduce a simple metric for SF. In logarithmic scale (right panel, Figure 1) oscillations are still present, but with smaller amplitudes. SFs (right panel in Figure 12) based on the combination of DRW and OpSim cadences show the oscillatory pattern in contrast to plain AGN light curves in g and r band obtained from OpSim1.5 DDF in Figure 15. These cases emphasize that the detection of binary candidates could be done via SFs.

Averaged SF deviations for OpSim1.5 (DDF, see Figure 13) are fragmented and noisy, indicating erratic behavior of SFs concerning homogenous SF. As much as the cadence is denser, the SFs resemble more those obtained from a homogenous light curve.

Some other essential factors could influence the relation between cadence, AGN variability observables, and their formal errors that we did not cover. Perhaps, the dependent and two independent variables used in our regression model are not the only critical light-curve characteristics that should be taken into the regression model. The more subtle influence will have light curves' nonstationarity, trends, and peak and valleys sharpness. For example the presence of a periodic signal in the light curve is imprinted in the periodic behavior of its cross-correlation function (CCF). Suppose the underlying signal in the light curve is complicated, such as

$$signal = t \sin(\theta t)$$

where t is time. In that case, the value of any local maximum of this function is greater than the values of all previous local maxima, and localization of the maxima of the CCF cannot be done using the highest value of the CCF correlation coefficient, which will bias the measurement of time lag.

There are some other conceptions that we did not cover in this study. Namely, it has been reported that the characteristics of SF for AGN light curves scales with physical parameters of AGN, for example, long time scale RMS variability SF^∞ is anti-correlated with AGN luminosity (MacLeod et al. 2010). However, its characteristic time scale is correlated with AGN luminosity (Sun et al. 2018) and short term RMS variability is sensitive to AGN bolometric luminosity (Sun et al. 2018). Thus it will be possible to characterize how much bias in the SF measurement is introduced by the LSST observing cadence. We plan to focus on such analysis in our future work when the Rubin Observatory releases Data Preview (DP) with AGN data. Based on three characteristic shape parameters of SF (SF^∞ , its associated characteristic time scale τ and power-law slope β of SF defined as $SF \sim (\frac{\Delta t}{\tau})^\beta$ the observed SF can be fitted with (see Sun et al. 2018):

$$SF(\Delta t|\tau, \sigma) = \sigma^2 \tau \left(1 - e^{(-\frac{\Delta t}{\tau})^\beta}\right) + \sigma_p \quad (17)$$

where σ_p is the uncertainty of the magnitude difference between two observations at time distance Δt . From the 'ideal' light curve, one can measure the 'reference' parameters: $Q = (SF^\infty, \tau, \beta)$ by fitting the observed SF using the above model. Then the shape parameters for the 'gappy' SFs obtained by the same fitting procedure for the reference SF can be compared with the reference shape parameters

$$\delta Q(i) = \frac{Q(i)^{ref} - Q(i)^{gappy}}{Q(i)^{ref}}. \quad (18)$$

5. Conclusion

To assess the observing strategies of the present and future spectroscopic and photometric surveys, we examined AGN variability-related observables (time lag, periodic oscillations, and SF) and their relation to predicting the most suitable LSST-like cadences. The first two observables affect constraining reliable AGN models and the third is essential for determination of the covariance function of the AGN light curves and can display oscillatory signals.

From this perspective, we constructed a multiple regression model to statistically identify the cadence-formal error pattern knowing AGN-variability observables from different surveys. We tested the performance of the proposed regression model on case studies of real (pre-LSST era) and an artificial observing strategies (LSST-like). Multiple regression model abstracts from the details of the real and artificial light curves, but establishes the general relationship with AGN-variability observables. We employed two different observing strategies: the optically uniform dataset including decade-long reverberation mapping campaigns of eight type 1 AGN, with distinct variability and optical spectra characteristics which is difficult to simulate; and the artificial data set, simulated according to the DRW method with an added periodic oscillation. The artificial data sets are constructed on several idealized and LSST-like observing strategies. For examination of cadence effects on SF we used only the idealized observing strategy. For time lag extraction we used the Gaussian process light curve modeling, and for the periodicity detection our 2DHybrid method. The results of our analysis are as follows:

- The two model versions (based on real and LSST-like observing strategy) predict comparable cadences for time lags and oscillation detection, whereas at $F_{var} \sim 10\%$ cadences for oscillation detection differ. The difference might be explained by sensitivity of oscillation detection to the light curve variability, because artificial set of objects contains only one object with $F_{var} \sim 10\%$. In general for time-lag and periodicity, for objects with higher ($\sim 20\%$) variability the predicted cadences are larger than those estimated for $F_{var} \sim 10\%$. As expected the predicted cadences for time-lag and periodicity are decreasing with assumed redshift of the object. The proposed multiple regression has shown promising potential for predicting AGN time-lag and periodicity cadences, but before such estimation can eventually meet observable practice, the regressions should be tested further in larger sets of object samples such as SDSS RM campaign.
- We find that, for SFs constructed on both idealized and LSST OpSim cadences, if the light curves contains periodic signal, the same oscillatory signal is seen in the large SFs time scales. We defined the simple metric to measure the properties of the SF, accounting for the deviation of the observed SF with respect to the ideal light curves. We showed that light curves with reasonable gaps would preserve the SF shape, and that even with larger gaps, some strategies of denser sampling could help to get smaller deviations of SF of "gappy" light curves from the SF constructed on homogeneous 1-day cadence.
- The smallest deviations of gapped SFs from idealized SF are observed when cadences are highly idealized or very dense LSST DDF cadences, having about 1500 observations in r filter. However, sparse DDF cadences in g filter indicate that gappy SFs would significantly deviate from homogeneous SF.
- We predict that the PDF of number of CB-SMBH, that LSST will detect on average during its lifetime, would have a peak at two dozens of objects for $q = 1$ and $q = 0.5$. However, the

PDF would peak at about 100 objects for $q = 0.05$. Based on constructed PDFs the cadences between ~ 20 and ~ 80 days are required for majority detection of binary candidates which is alike to multiple regression model prediction.

The multiple regression model presented may be used in assessing observing strategies of the present and future photometric and spectroscopic surveys, such as the LSST, MSE, SDSS-V, and many other to come. For the purpose of granting scientists to easily review and analyse the method described in this paper, we have developed various Jupyter notebooks. Our code is publicly available as open-source code on GitHub (https://github.com/LSST-sersag/agn_cadences).

Acknowledgements. The authors acknowledge funding provided by the Faculty of Mathematics University of Belgrade (the contract 451-03-9/2021-14/200104), Astronomical Observatory (the contract 451-03-68/2020-14/200002) and Faculty of Science University of Kragujevac (the contract 451-03-9/2021-14/200122), through the grants by the Ministry of Education, Science, and Technological Development of the Republic of Serbia.

References

- Barth A. J., Pancoast A., Thorman S. J. et al. 2011, *ApJ*, 743, 4
 Barth A. J., Pancoast A., Bennert V. N. et al. 2013, *ApJ*, 769, 128
 Barth A. J., Bennert V. N., Canalizo G. et al. 2015, *ApJS*, 217, 26
 Bentz, M. C., Denney, K. D., Grier, C. J., Barth, A. J., Peterson, B. M., Vestergaard, M. et al. 2013, *ApJ*, 767, 149
 Bentz M. C., Walsh J. L., Barth A. J. et al. 2008, *ApJL*, 689, L21
 Bentz M. C., Walsh J. L., Barth A. J. et al. 2009, *ApJ*, 705, 199
 Bentz M. C., Peterson, B. M., Netzer, H., Pogge, R. W., Vestergaard, M. 2009, *ApJ*, 697, 160
 Biswas, R., Daniel, S.F., Hložek, R., Kim, A.G. et al. 2020, *ApJS*, 247, 60
 Bon, E., Jovanović, P., Marziani, P., Shapovalova, A. I., Bon, N., Borka Jovanović, V. et al. 2012, *ApJ*, 759, 2, id.118
 Bon, E., Zucker, S., Netzer, H. et al. 2016, *ApJS*, 225, 29
 Brandt, W. N., Ni, Q., Yang, G., Anderson, S. F., Assef, R. J., Barth, A. J. et al. 2018, white paper on LSST cadence optimization, arXiv:1811.06542
 Burke-Spolaor, S., Taylor, S.R., Charisi, M. et al. 2019 *Astron Astrophys Rev* 27, 5
 Caplar, N., Lilly S., J., Trakhtenbrot, B. 2017, *ApJ*, 834, 111
 Chelouche, D., Daniel, E. 2012, *ApJ*, 747, 62
 Chen, Y. C. 2017, *Biostatistics & Epidemiology*, 1, 161
 Charisi, M., Haiman, Z., Schiminovich, D., D' Orazio, D. J. 2018, *MNRAS*, 476, 4617
 Denney K. D., Peterson B. M., Pogge R. W. et al. 2009, *ApJL*, 704, L80
 Denney, K. D., De Rosa, G., Croxall, K., Gupta, A., Bentz, M. C. et al. 2014, *ApJ*, 796, id. 134
 D'Orazio, D. J., Haiman, Z., MacFadyen, A. 2013, *MNRAS*, 436, 2997
 D' Orazio, D. J., Haiman, Z., Schiminovich, D. 2015, *Nature*, 525, 351
 D' Orazio D. J., Haiman Z. 2017, *MNRAS*, 470, 1198
 D' Orazio, D., Loeb, A. 2019, *Phys. Rev. D.*, 100, 103016
 Du P., Hu C., Lu K.-X., Wang, F., Qui, J., Li, Y.-R., Bai, J.-M., Kaspi, S., Netzer, H., Wang, J.-M. 2014, *ApJ*, 782, 45
 Du P., Hu C., Lu K.-X., Huang, Y.-K., Cheng, C., Qiu, J., Li, Y.-R. et al. 2015, *ApJ*, 806, 22
 Du P., Lu K.-X., Zhang Z.-X., Huang, Y.-K., Wang, K., Hu, C., Qiu, J., Li, Y.-R. et al. 2016, *ApJ*, 825, 126
 Du P., Zhang Z.-X., Wang K., Huang, Y.-K., Zhang, Y., Lu, K.-X., Hu, C., Li, Y.-R. et al. 2018, *ApJ*, 856, 6
 Edelson, R., Gelbord, J., Cackett, E., Peterson, B. M., Horne, K., Barth, A. J. et al. *ApJ*, 870, 2, id. 123
 Elvis, M., 2001, in *The Century of Space Science* (eds J.A. Bleeker, J. Geiss, and M. Huber), Kluwer Academic Publishers, 529
 Farris, B. D., Duffell, P., MacFadyen, A. I., Haiman, Z. 2014, *ApJ*, 783, 134
 Faisst, A. L., Prakash, A., Capak, P. et al. 2019, *ApJL*, 881, L9
 Graham, M. J., Djorgovski, S. G., Stern, D., Glikman, E., Drake, A. J. et al. 2015, *Nature*, 518, 74
 Graham, M. L., Connolly, A.J., Ivezić, Ž., Schmidt, S. J., Lynne Jones, R., Jurić, M. et al. 2018, *AJ*, 155, 1
 Grier C. J., Peterson B. M., Pogge R. W. et al. 2012, *ApJ*, 755, 60
 Grier C. J., Trump J. R., Shen Y. et al. 2017, *ApJ*, 851, 21
 Hopkins, P. F., Richards, G. T., Hernquist, L. 2007, *ApJ*, 654, 731
 Horne, K., Peterson, B. M., Collier, S. J., Netzer, H. 2004, *PASP*, 116, 819, 465

Ilić, D., Oknyansky, V., Popović, L. Č., Tsygankov, S. S., Belinski, A. A., Tatarnikov, A. M., Dodin, A. V., Shatsky, N. I., Ikonnikova, N. P., Rakić, N., Kovačević, A., Marčeta-Mandić, S. et al. 2020, *A&A*, 638, A13

Ivezić, Ž., Kahn, S. M., Tyson, A., Abel, B., Acosta, E. et al. 2019, *ApJ*, 873, 44

Jones, L., Yoachim, P., Ivezić, v Z., Jurić, M., Eggl, S., Chesley, S., Fraser, W., et al. 2020, *DPS*

Jun, H. D., Stern, D., Graham, M. J., Djorgovski, S. G., Mainzer, A., Cutri, R., M., Drake, A. J., Mahabal, A. A. 2015, *ApJ*, 814, L12

Kelly, B. C., Bechtold, J., Siemiginowska, A. 2009, *ApJ*, 698, 895

Kelly, B. C., Treu, T., Malkan, M., Pancoast, A., Woo, J.-H. 2013, *ApJ*, 779, 187

Kim, J., Im, M., Choi, C., Hwang, S. 2019, *ApJ*, 884, 103

Kollmeier, J. A., Onken, C. A., Kochanek, C. S., Gould, A. et al. 2006, *ApJ*, 648, 128

Kollatschny, W., Fricke, K. J. 1985, *A&A*, 146, L11

Kovačević, A. B., Pérez-Hernández, E., Popović, L. Č., Shapovalova, A. I., Kollatschny, W., Ilić, D. 2018, *MNRAS*, 475, 2051

Kovačević, A., Popović, L. Č., Shapovalova, A. I., Ilić, D. 2017, *Ap& SS*, 362, id. 31

Kovačević, A., Popović, L. Č., Ilić, D. 2020, *Open astronomy*, 29, 51

Kovačević, A. B., Popović, L. Č., Simić, S., Ilić, D. 2019, *ApJ*, 871, article id. 32

Kovačević, A. B., Yi, T., Dai, X., Yang, X., Čvorović-Hajdinjak, I., Popović, L. Č. 2020, *MNRAS*, 494, 4069

Kozłowski, S. 2016, *ApJ*, 826, 118

Li, Y.-R., Wang, J.-M., Ho, L. C., Lu, K.-X., Qiu, J., Du, P., Hu, C., Huang, Y.-K., Zhang, Z.-X., Wang, K., Bai, J.-M. 2016, *ApJ*, 822, 4

Liu, T., Gezari, S., Coleman Miller, M. 2018, *ApJ*, 859, id. L12

Lyutyj, V. M., Oknyanskij, V. L., Chuvaev, K. K. 1984, *Soviet Astronomy Letters*, vol. 10, 335

MacLeod, C. L., Ivezić, Ž., Kochanek, C. S., Kozłowski, S., Kelly, B. et al. 2010, *ApJ*, 721, 1014

Marshall, P., Anguita, T., Bianco, F., Bellm, E.C., Brandt, W. N. et al. 2017, preprint, arXiv:1708.04058

Moreno, J., Vogeley, M. S., Richards, G. T., Yu, W. 2019, *PASP*, 131, 063001.

The MSE Science Team, Babusiaux, C., Bergemann, M., et al. 2019, arXiv:1904.04907

Neira F., Anguita, T., Vernardos, G. 2020, *MNRAS*, 495, 544.

Núñez Pozo, F., Haas, M., Ramolla, M., Bruckmann, C. et al. 2014, *A&A*, 568, A36

O'Brien P.T., Dietrich, M., Leighly, K et al. 1998, *ApJ*, 509, 163

Peterson B. M., Wanders I., Bertram R. et al. 1998, *ApJ*, 501, 82

Peterson B. M., Berlind P., Bertram R. et al. 2002, *ApJ*, 581, 197

Peterson, B. M., Ferrarese, L., Gilbert, K. M., Kaspi, S. et al. 2004, *ApJ*, 613, 682

Popović, L. Č., Shapovalova, A. I., Ilić, D., Kovačević, A., Kollatschny, W. et al. 2011, *A&A*, 528, A130

Popović, L. Č., Shapovalova, A. I., Ilić, D., Burenkov, A. N., Chavushyan, V. H.; Kollatschny, W., Kovačević, A. et al. 2014, *A&A*, 572, A66

Shablovinskaya, E. S., Afanasiev, V. L., Popović, L. Č. 2020, *ApJ*, 892, id.118

Shankar, F., Weinberg, D. H., Miralda-Escude, J. 2009, *ApJ*, 690, 20

Shapovalova, A. I., Burenkov, A. N., Carrasco, L., Chavushyan, V. H. et al. 2001, *A&A*, 376, p. 775

Shapovalova, A. I., Doroshenko, V. T., Bochkarev, N. G., Burenkov, A. N. et al. 2004, *A&A*, 422, 92

Shapovalova, A. I., Popović L. Č., Collin, S., Burenkov, A. N. et al. 2008, *A&A*, 486, 99

Shapovalova, A. I., Popović, L. Č., Burenkov, A. N., Chavushyan, V. H., Ilić, D., Kovačević, A. et al. 2010a, *A&A*, 509, A106

Shapovalova, A. I., Popović, L. Č., Burenkov, A. N., Chavushyan, V. H., Ilić, D., Kollatschny, W., Kovačević, A. et al. 2010b, *A&A*, 517, A42

Shapovalova, A. I., Popović, L. Č., Burenkov, A. N., Chavushyan, V. H., Ilić, D., Kovačević, A., Kollatschny, W., Kovačević, J. et al. 2012, *ApJS*, 202, 10

Shapovalova, A. I., Popović, L. Č., Burenkov, A. N.; Chavushyan, V. H., Ilić, D., Kollatschny, W., Kovačević, A. et al. 2013, *A&A*, 559, A10

Shapovalova, A. I., Popović, L. Č., Chavushyan, V. H., Burenkov, A. N., Ilić, D., Kollatschny, W., Kovačević, A. et al. 2016, *ApJS*, 222, id. 25

Shapovalova, Alla I, Popović, L. Č., Chavushyan, V. H., Afanasiev, V. L., Ilić, D., Kovačević, A. et al. 2017, *MNRAS*, 466, 4759

Shapovalova, A. I., Popović, L. Č., Afanasiev, V. L., Ilić, D., Kovačević, A., Burenkov, A. N., Chavushyan, V. H., Marčeta-Mandić, S. et al. 2019, *MNRAS*, 485, 4790

Shen, Y., Anderson, S., Berger, E. et al. 2019, *BASS*, 51, 274

Shen Y., Horne K., Grier C. J., Peterson, B. M., Denney, K. D. et al. 2016, *ApJ*, 818, 30

Shen, Y., Hall, P. B., Horne, K., et al. 2019, *ApJS*, 241, 34

Shi, J.-M., Krolik, J. H. 2015, *ApJ*, 807, 131

Simonetti, J. H., Cordes, J. M., Heeschen, D. S. 1985, *ApJ*, 296, 46

Sun, M., Xue, Y., Wang, J., Cai, Z., Guo, H. 2018, *ApJ*, 866, 74

Wang, J.-M., Du, P., Hu, C., Hu, C., Netzer, H. et al. 2014, *ApJ*, 793, 108

Wang, J., Songsheng, Y., Li, Y. R., Du P., Zhang, Z. X. 2020, *Nat. Astron.*, 4, 517

Wang, J. M., Xu, D., W., Wei, J. Y. 2018, *ApJ*, 858, 49

Wang, H., Yin, C., Xiang, F. 2017, *ApSS*, 362, 99

Woo, J. H., Urry, C. M. 2002, *ApJ*, 579, 530

Vaughan, S. 2010, *MNRAS*, 402, 307

Vaughan, S., Uttley, P., Markowitz, A. G., Huppenkothen, D. et al. 2016, *MNRAS*, 461, 3145

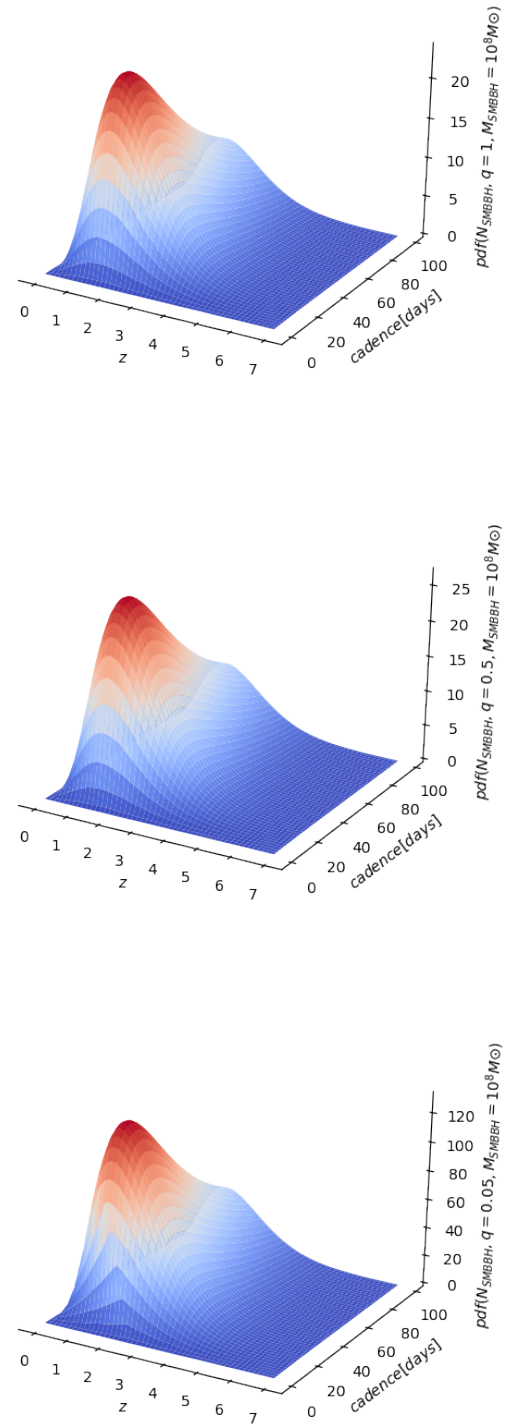


Fig. 14. Probability density functions of expected number of CB-SMBH with total mass $M_{tot} = 10^8 M_\odot$ and mass ratios $q = 1, 0.5, 0.05$ (from left to right respectively), as the probability for finding an CB-SMBH at the orbital period resolvable by LSST.

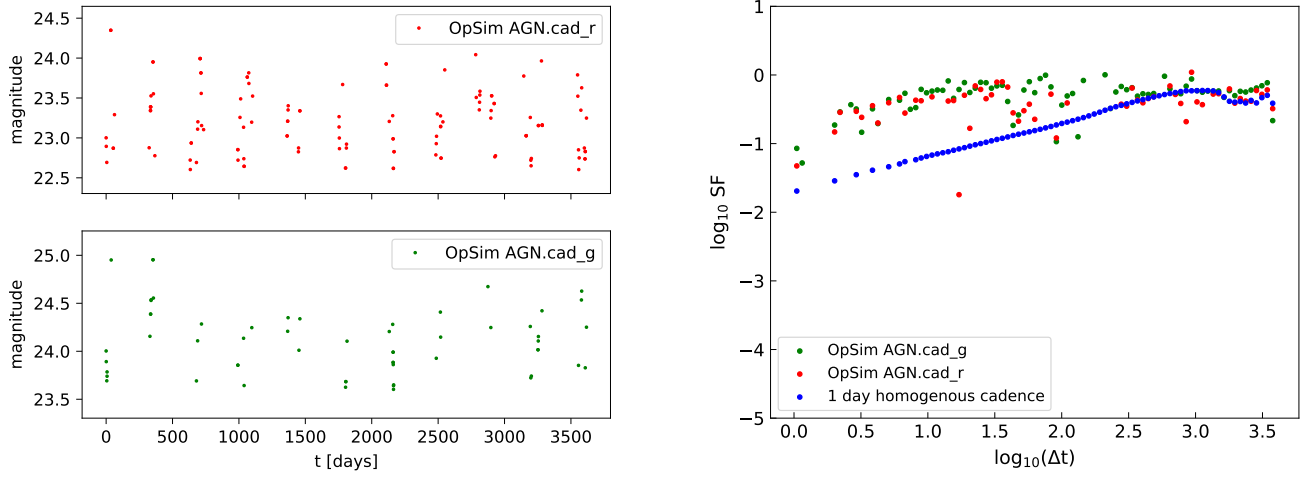


Fig. 15. *Left:* AGN Light curve in r and g band obtained from OpSim. *Right:* Corresponding SFs compared to the SF of homogeneous 1-day cadence light curve.