The Astrophysical Journal, 826:118 (16pp), 2016 August 1 doi:10.3847/0004-637X/826/2/118

REVISITING STOCHASTIC VARIABILITY OF AGNs WITH STRUCTURE FUNCTIONS

Szymon Kozłowski
Warsaw University Observatory, Al. Ujazdowskie 4, 00-478 Warszawa, Poland; simkoz@astrouw.edu.pl
Received 2016 March 12; revised 2016 May 12; accepted 2016 May 16; published 2016 July 26

ABSTRACT

Discrepancies between reported structure function (SF) slopes and their overall flatness as compared to the expectations from the damped random walk (DRW) model, which generally well describes the variability of active galactic nuclei (AGNs), have triggered us to study this problem in detail. We review common AGN variability observables and identify their most common problems. Equipped with this knowledge, we study ∼9000 *r*-band AGN light curves from Stripe 82 of the Sloan Digital Sky Survey, using SFs described by stochastic processes with the power exponential covariance matrix of the signal. We model the "subensemble" SFs in the redshift–absolute magnitude bins with the full SF equation (including the turnover and the noise part) and a single power law (SPL; in the "red noise regime" after subtracting the noise term). The distribution of full-equation SF (SPL) slopes peaks at $\gamma = 0.55 \pm 0.08$ (0.52 ± 0.06) and is consistent with the DRW model. There is a hint of a weak correlation of γ with the luminosity and a lack of correlation with the black hole mass. The typical decorrelation timescale in the optical is $\tau = 0.97 \pm 0.46$ year. The SF amplitude at one year obtained from the SPL fitting is $\mathrm{SF}_0 = 0.22 \pm 0.06$ mag and is overestimated because the SF is already at the turnover part, so the true value is $\mathrm{SF}_0 = 0.20 \pm 0.06$ mag. The asymptotic variability is $\mathrm{SF}_\infty = 0.25 \pm 0.06$ mag. It is strongly anticorrelated with both the luminosity and the Eddington ratio and is correlated with the black hole mass. The reliability of these results is fortified with Monte Carlo simulations.

Key words: accretion, accretion disks – galaxies: active – methods: data analysis – quasars: general

1. INTRODUCTION

One of the flagship properties of active galactic nuclei (AGNs) is their variability. It was discovered (Smith & Hoffleit 1963; Matthews & Sandage 1963) right after or along with the discovery of this class of objects in 1963 (Greenstein 1963; Schmidt 1963). Because AGNs are the most powerful sources of continuous light in the universe (with the bolometric luminosities as high as 10^{48} erg s^{-1} or $10^{14} L_\odot$), their variability at timescales of months to years of the order of 10% of the total light must be enormous. There are multiple lines of evidence that AGNs are powered by the accretion of matter onto supermassive black holes (SMBHs; Shakura & Sunyaev 1973; Rees 1984). For example, chromatic microlensing of strongly gravitationally lensed AGNs provides us with the sizes (∼1 au in optical) and temperature profiles of accretion disks (Czerny et al. 1994; Kochanek 2004; Dai et al. 2010; Morgan et al. 2010; Blackburne et al. 2011).

During the five decades since their discovery, we have witnessed an increasing number of studies on the variability of AGNs, on both data analyses and theory. The theoretical studies examined accretion disk instabilities, surface temperature fluctuations, and variable heating from coronal X-rays (e.g., Shakura & Sunyaev 1976; Rokaki et al. 1993; Chen & Taam 1995; Kawaguchi et al. 1998; Ruan et al. 2014) but also currently considered as nonviable: microlensing and chain supernovae, also known as the starburst model (Hawkins 1993; Baganoff & Malkan 1995; Kawaguchi et al. 1998). On the data side, aperiodic luminosity fluctuations have been described by means of the rms variability (sometimes the squared rms) as a function of the time difference, or equivalently frequency, between epochs, and these methods are known as the structure function (SF) and power spectral density (PSD) analyses, respectively. Continuously growing, in length and cadence, quasar light curves from optical surveys enabled even more detailed studies and quantification of AGN variability that include direct light curve modeling. A model that works particularly well is the damped random walk model (DRW; Kelly et al. 2009; Kozłowski et al. 2010a; MacLeod et al. 2010; Butler & Bloom 2011; MacLeod et al. 2011; Zu et al. 2011; MacLeod et al. 2012; Ruan et al. 2012; Zu et al. 2013a), although various alternatives have been tested (e.g., Kelly et al. 2011; Zu et al. 2013a; Pancoast et al. 2014), and some apparent deviations from it have been reported (Mushotzky et al. 2011; Kasliwal et al. 2015a). DRW is also the simplest of a broader, more general class of continuous-time autoregressive moving average (CARMA) models, recently considered by Kelly et al. (2014) and Simm et al. (2016).

The first goal of this paper is to *explicitly* review AGN variability observables, in particular their connection to the core of the underlying process. We will be primarily interested in the SF analysis, but we also skim the PSD analysis, both typically considered for sparse or short light curves. We briefly review the DRW method (or its siblings, e.g., Zu et al. 2013a) used to model light curves (Figure 1). The importance of the first two of these observables lies in their model-independent capability to directly measure the covariance function of the signal (i.e., pinpoint the underlying process), while the DRW light curve modeling lacks this capability (Kozłowski 2016). The second goal is to apply the SF method to ∼9000 quasar light curves from the Sloan Digital Sky Survey (SDSS) Stripe 82 (S82) in order to find the covariance function of the process and also basic variability parameters. Subsequently we will be interested in correlations between the shape of the covariance function and the physical parameters of AGNs, such as the black hole mass, luminosity, Eddington ratio, and redshift (or equivalently the rest-frame wavelength). By means of Monte Carlo simulations, we will show how the SF or DRW is affected by data quality or sampling; in particular we present the biases in the derived SF and DRW parameters.

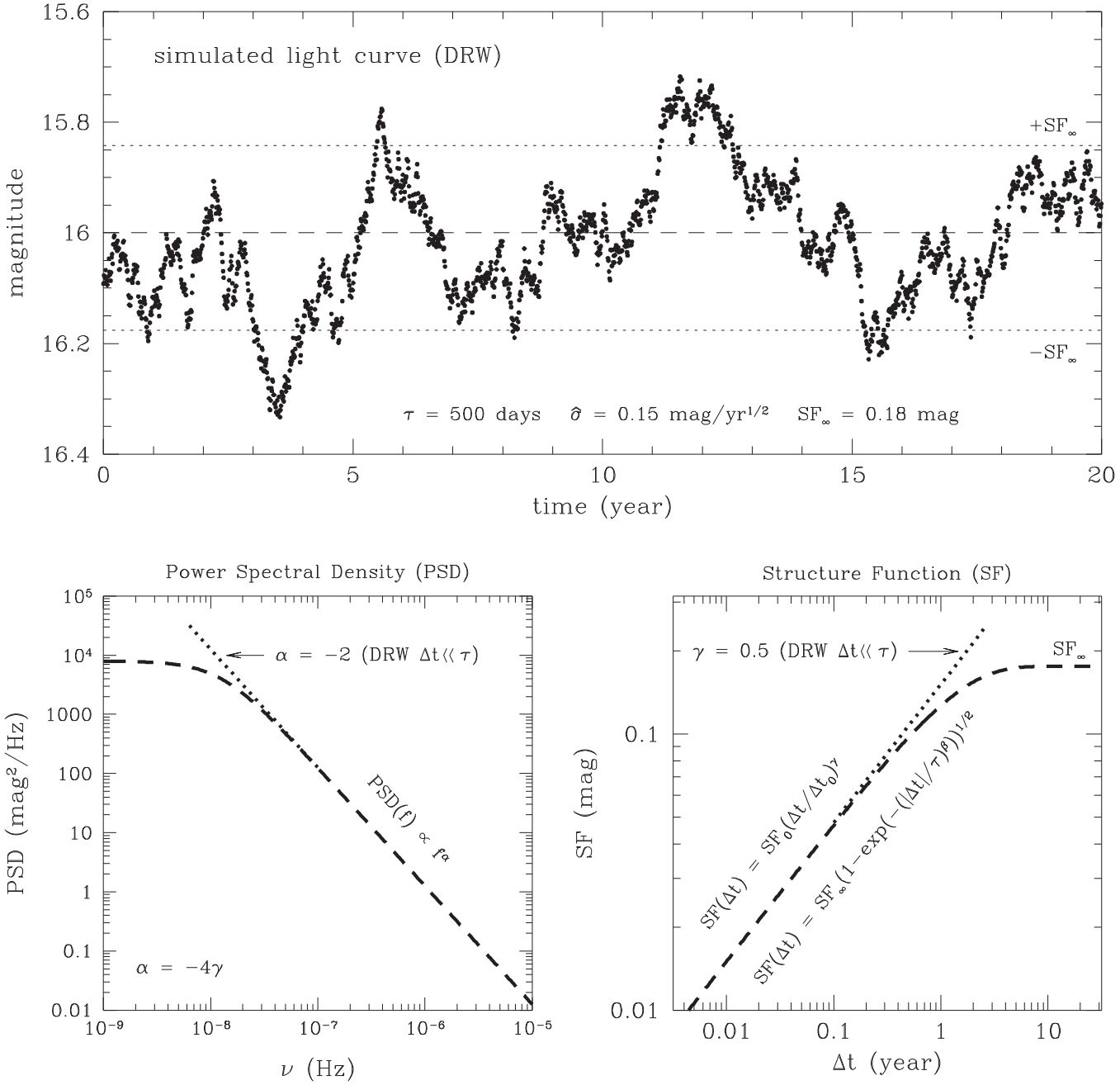

Figure 1. Presentation of basic AGN variability concepts and measures. Top panel: an idealized simulated AGN light curve using DRW is shown. It can be either directly fit by the DRW (or CARMA) model and return the model parameters or studied via the PSD (bottom left panel) or SF (bottom right panel) analysis. Each panel also presents basic variability features related to these AGN variability measures.

The paper is organized as follows. In Section 2, we describe and discuss variability observables for AGNs. In Section 3, we analyze the SDSS S82 data with the SF and DRW models and simulate similar data sets in order to understand any problems, biases, or systematics. In Section 4 we discuss our findings. The paper is summarized in Section 5.

2. DESCRIPTION OF VARIABILITY

We will now explicitly review vital details of measuring AGN variability. We briefly recollect that AGN variability resembles and is well described by stochastic processes: it is aperiodic, meaning it has smooth PSDs (with no peaks on periodograms; e.g., Kelly et al. 2009). Throughout this paper we will explicitly assume that AGN variability is due to a single stochastic process or a combination of stochastic processes (Kelly et al. 2011).

2.1. SF in a Nutshell

Let us consider a data set composed of a collection of measured data y_i (e.g., magnitudes) at times t_i with $i = 1,...,N$ points (hereafter a light curve) that can be represented as a sum of the true signal s_i and noise n_i: $y_i = s_i + n_i$ (see Scargle 1981, 1982, 1989; Press et al. 1992a, 1992b; Rybicki & Press 1992 for an intuitive description of a time series analysis). Having an AGN light curve, we will be interested in the relationship of this light curve to its copy shifted by time Δt

for various $\Delta t = (0, \ldots, \infty)$ days, because, as we will show, such a relationship contains the key to understanding the origin of variability. For each time shift Δt, often called the time "lag," we will find all pairs of points occurring at the same instant with $\Delta t = |t_i - t_j|$, where i and j are the indices in the original and shifted light curves. Because AGNs are typically distant sources, they participate in the Hubble flow, and their redshifts z are often substantial. We correct the observed-frame variability to the rest frame via $\Delta t = |t_i - t_j|(1+z)^{-1}$. Throughout this paper we will assume Δt to be the rest-frame time lag.

If an underlying stochastic process leading to variability is stationary (its mean, variance, and probability do not change with time), then the relationship between the two light curves (the original and the shifted one) can be quantified by the covariance function

$$\begin{aligned}\mathrm{cov}(\Delta t) &\equiv \mathrm{cov}(y(t), y(t+\Delta t)) \\ &\equiv \mathrm{cov}(y_i, y_j) \equiv \\ &\equiv \frac{1}{N_{\Delta t\ \mathrm{pairs}}} \sum_{i=1}^{N_{\Delta t\ \mathrm{pairs}}} (y_i - \langle y\rangle)(y_j - \langle y\rangle) \\ &\equiv \langle (y_i - \langle y\rangle)(y_j - \langle y\rangle)\rangle, \end{aligned} \quad (1)$$

and the covariance for $\Delta t = 0$ days ($y_i = y_j$) is

$$\mathrm{cov}(0) \equiv \mathrm{cov}(y_i, y_i) \equiv \mathrm{var}(y_i) \equiv \langle (y_i - \langle y\rangle)^2\rangle, \quad (2)$$

where var (y_i) is the data variance and $\langle y\rangle$ is the mean.

It is straightforward to show (Press et al. 1992a) that the covariance of data (from Equation (1)) can be expressed in terms of the data variance and the SF as

$$\mathrm{cov}(y_i, y_j) \equiv \mathrm{var}(y_i) - V(y_i, y_j), \quad (3)$$

where

$$V(y_i, y_j) = \frac{1}{2}\langle (y_i - y_j)^2\rangle \quad (4)$$

is the "theoretical" SF (in units of squared magnitude), as opposed to the typically reported one that scales as $\mathrm{SF} = \sqrt{2V}$ (in units of magnitude).

Throughout this paper, we will be interested in the form of the covariance function of the signal cov (s_i, s_j) (or equivalently in the autocorrelation function (ACF), which by definition is $\mathrm{ACF}(\Delta t) \equiv \mathrm{cov}(s_i, s_j)/\sigma_s^2$) describing the underlying process leading to variability. It can be directly obtained from the data through the structure function (from Equation (3)):

$$\begin{aligned} V(y_i, y_j) &= \mathrm{var}(y_i) - \mathrm{cov}(y_i, y_j) = \mathrm{var}(s_i) \\ &\quad + \mathrm{var}(n_i) - \mathrm{cov}(s_i, s_j) - \mathrm{cov}(n_i, n_j) \\ &= \sigma_s^2 + \sigma_n^2 - \mathrm{cov}(s_i, s_j), \end{aligned} \quad (5)$$

where $\mathrm{var}(s_i) \equiv \sigma_s^2$, $\mathrm{var}(n_i) \equiv \sigma_n^2$, and $\mathrm{cov}(s_i, n_i) = \mathrm{cov}(n_i, n_j) \equiv 0$ because both the data and noise are assumed to be uncorrelated with the noise. We implicitly assumed above that both the signal and noise have Gaussian distributions (var $\equiv \sigma^2$); however, while we will be interested in Gaussian processes for the signal, the noise may have a non-Gaussian distribution ($\mathrm{var}(n_i) \neq \sigma_n^2$). From Equation (5) it follows that the rest-frame SF is

$$\mathrm{SF}_{\mathrm{obs}}(\Delta t) = \sqrt{2\sigma_s^2 + 2\sigma_n^2 - 2\mathrm{cov}(s_i, s_j)} \quad (6)$$

and contains information on the covariance function of the process causing the variability. The true underlying SF, after subtracting the noise term $2\sigma_n^2$ or $2\mathrm{var}(n)$ for a non-Gaussian distribution, is

$$\mathrm{SF}_{\mathrm{true}}(\Delta t) = \sqrt{\mathrm{SF}_{\mathrm{obs}}(\Delta t)^2 - 2\sigma_n^2}, \quad (7)$$

and hence

$$\begin{aligned}\mathrm{SF}_{\mathrm{true}}(\Delta t) &= \sqrt{2\sigma_s^2(1 - \mathrm{ACF}(\Delta t))} \\ &= \mathrm{SF}_\infty\sqrt{1 - \mathrm{ACF}(\Delta t)}, \end{aligned} \quad (8)$$

and MacLeod et al. (2010) defines $\mathrm{SF}_\infty = \sqrt{2}\,\sigma_s$ because $\mathrm{ACF}(\Delta t) \to 0$ as $\Delta t \to \infty$. The ACF can take any value between 0 and 1, and its shape is simply a function of Δt (Figure 2). It can be also thought of as a "memory function," where ACF = 1 (ACF = 0) reflects the perfect (lack of) memory, and any values between 0 and 1 show how strongly two data points separated by Δt "remember" each other.

The calculation of $\mathrm{SF}_{\mathrm{obs}}$ is typically obtained from (Equation (4)):

$$\begin{aligned}\mathrm{SF}_{\mathrm{obs}}(\Delta t) &= \sqrt{\frac{1}{N_{\Delta t\ \mathrm{pairs}}} \sum_{i=1}^{N_{\Delta t\ \mathrm{pairs}}} (y(t) - y(t+\Delta t))^2} \\ &\equiv \mathrm{rms}[y(t) - y(t+\Delta t)] \end{aligned} \quad (9)$$

or equivalently

$$\mathrm{SF}_{\mathrm{obs}}(\Delta t) = 0.741 \times \mathrm{IQR}, \quad (10)$$

where IQR is the interquartile range between 25% and 75% of the sorted $(y(t) - y(t+\Delta t))$ distribution, and the 0.741 coefficient is the conversion to σ for the normal or Gaussian distribution (MacLeod et al. 2012). The latter description provides the rms values less affected by outliers in the distribution or if the distribution is non-Gaussian. From Equation (9) it is obvious that the SF measures the amount of rms variability as a function of the time interval or "lag" (Δt) between points.

Note also that the rms around the mean is $\mathrm{rms}^2[x - \langle x\rangle] = \langle x\rangle^2 + \sigma_x^2$, but in our case $\langle x\rangle = 0$. We also calculate the rms of the differences (and not around the mean), so $\mathrm{rms}^2[x_i - x_j] = 2\sigma_x^2$. Equation (7) becomes then

$$\begin{aligned}\mathrm{SF}^2_{\mathrm{true}}(\Delta t) \equiv\ & \mathrm{rms}^2[y(t) - y(t+\Delta t)] \\ & - \mathrm{rms}^2[y(t) - y(t + (\Delta t \to 0))] \end{aligned}, \quad (11)$$

and a variant of it was used to study the SFs of *Spitzer* mid-infrared AGNs in Kozłowski et al. (2010b, 2016) (the second term was the rms for the nonvariable field objects without the requirement $\Delta t \to 0$).

In fact, Press et al. (1992a) seems to provide the most correct way to estimate SFs (although sensitive to outliers) from

$$\begin{aligned}\mathrm{SF}^2_{\mathrm{true}}(\Delta t) = & \frac{1}{N_{\Delta t\ \mathrm{pairs}}} \sum_{i=1}^{N_{\Delta t\ \mathrm{pairs}}} (y(t) - y(t+\Delta t))^2 \\ & - \sigma_n^2(y(t)) - \sigma_n^2(y(t+\Delta t)), \end{aligned} \quad (12)$$

where each magnitude measurement $y(t)$ is accompanied by its own $\sigma_n(y)$ (i.e., dispersion of noise for this particular

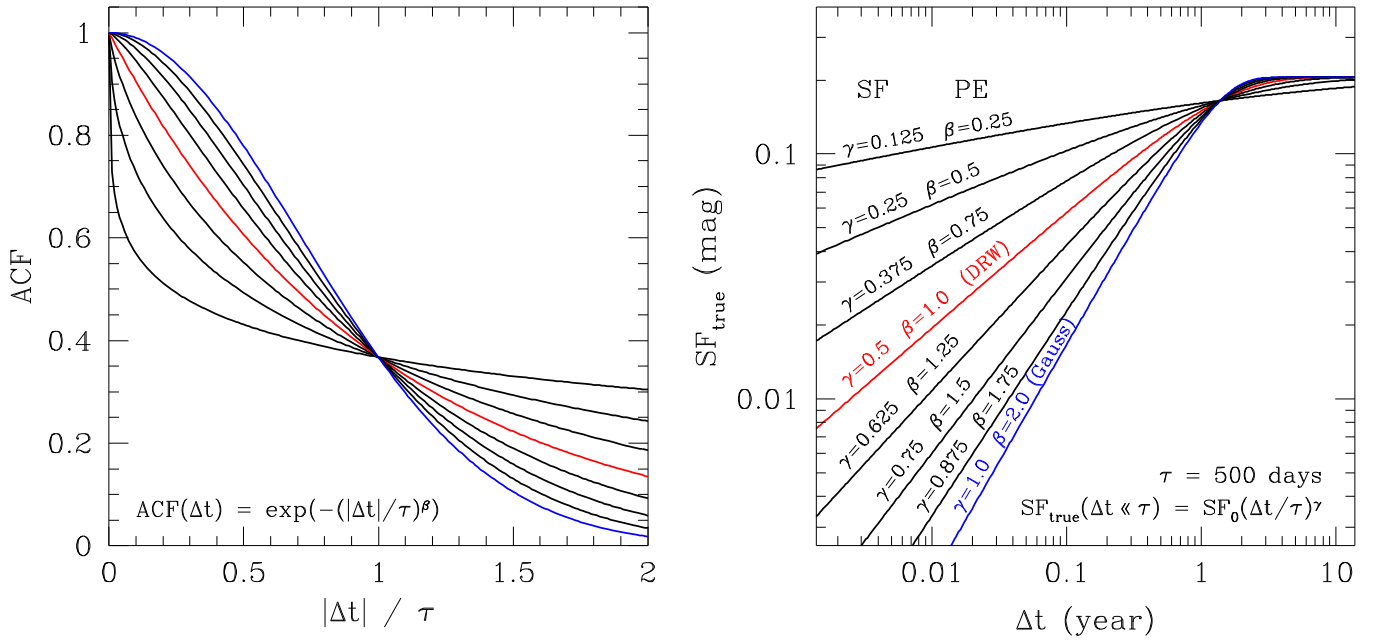

Figure 2. Graphic explanation of the connection between the power exponential ACFs (left panel) and their corresponding SFs (right panel). DRW (marked in red) is the case with $\beta = 1$ and produces the SF slope $\gamma = 0.5$ for $\Delta t \ll \tau$ (i.e., the red noise regime). Steeper or shallower SFs than DRW may be due to a combination of DRW processes and can be viewed as having a stronger ($\beta > 1$) or weaker ($\beta < 1$) ACF, respectively.

magnitude, typically estimated via dispersions from light curves of nonvariable field objects of the same magnitude) and must be obtained prior to the SF calculation.

2.2. The Photometric Noise

From Equation (6) it follows that the shape of the observed SF as a function of Δt strongly depends on σ_n for $\Delta t \to 0$ and is a function of a survey's photometric properties and a source magnitude. The noise term $2\sigma_n^2$ in Equation (7) could be directly calculated from the photometric error bars by assuming that $G(\sigma_\epsilon) = G(\sigma_n)$, which typically is not true as the measurements' uncertainties are often incorrectly estimated (see a discussion by, e.g., Skowron et al. 2016). There exist, operationally, better ways to do it: (1) it can be calculated as the rms of measurements (or 0.741 × IQR) for nonvariable field objects with the same magnitude (e.g., Kozłowski et al. 2010b, 2016), or (2) it can be directly estimated from $SF_{obs}(\Delta t)$ for $\Delta t \to 0$, because only then $SF_{true}(\Delta t) \to 0$. Please note that, by setting var $(n) = \sigma_n^2$, we implicitly assume the noise distribution to be Gaussian, but from Figure 3 we see that the SDSS S82 AGNs have a weakly non-Gaussian distribution of uncertainties. If the distribution of uncertainties can be approximated by any function for which the variance exists, instead of subtracting the noise term $2\sigma_n^2$, one should subtract 2var(n).

One may ask if the exponential wings of the distributions shown in Figure 3 are due to the photometric noise or rather are due to the underlying AGN variability. As already explained, $SF_{obs}(\Delta t) \approx 2\text{var}(n)$ for $\Delta t \to 0$ because only then $SF_{true}(\Delta t) \to 0$. We have checked the corresponding distributions for the field sources (stars, galaxies) as in Figure 3, and they show similar exponential wings, so they are not due to AGN variability.

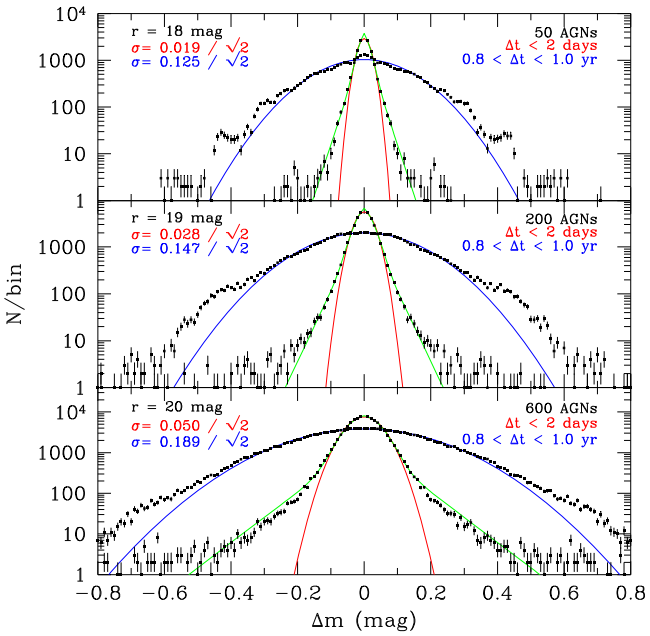

Figure 3. Distributions of magnitude differences Δm for two Δt regimes, $\Delta t < 2$ days (points with red or green fits) and $0.8 < \Delta t < 1$ year (points with blue fits), and for the AGN mean magnitudes falling into three levels, $r = 18 \pm 0.1$ mag, 19 ± 0.05 mag, and 20 ± 0.05 mag (panels from top to bottom). For $\Delta t < 2$ days, $SF_{true} \approx 0$ mag, so these distributions, in fact, show true uncertainty distributions (they need to be divided by $\sqrt{2}$) for a given magnitude. The photometric error bars are weakly non-Gaussian (red Gaussian fits) and are better described as a sum of a Gaussian and an exponential function (green fits). The distribution of differences at the longer lag is also weakly non-Gaussian (blue Gaussian fits).

2.3. Common SF Issues

One of the triggers for this study were both seemingly flat SFs ($\gamma = 0.1$–0.4, as compared to that required by DRW, $\gamma = 0.5$) and the discrepancies between them.

It is obvious that SFs (or PSDs, Kelly et al. 2011) cannot be single power laws (SPL) for $\Delta t \rightarrow \infty$ because a very large or infinite power would be required at long timescales. Therefore a typical SF is rather bent (Figure 2), and it has two (or sometimes three) regimes: for $\Delta t \ll \tau$ (the red noise regime) it can be fit as an SPL of the form

$$\mathrm{SF}(\Delta t) = \mathrm{SF}_0 \left(\frac{\Delta t}{\Delta t_0} \right)^{\gamma}, \tag{13}$$

where SF_0 is the variability amplitude at a fixed timescale Δt_0. Around $\Delta t \approx \tau$ we observe a transition, a turnover, from an SPL into another SPL with $\gamma = 0$ (i.e., the white noise), which is the second regime for $\Delta t \rightarrow \infty$. The third regime is for $\Delta t \approx 0$ days; in particular if the noise term is not subtracted (or subtracted incompletely), then we are in an another SPL regime with $\gamma = 0$ (i.e., white noise regime due to the photometric noise). This very short timescale regime vanishes provided the photometric noise is removed correctly. Because the real data rarely show a Gaussian-like behavior in the noise, SFs are uncertain in this regime, and it is best to simply exclude them from the analyses.

In early studies, the turnover timescale τ was not known (which is about 500 days in the optical, MacLeod et al. 2010, 2012, or about one year from the SFs in this study), and the SFs were fit as an SPL, which is obviously incorrect in light of the infinite power required to generate variations at long lags, and henceforth the measured SFs must have been flat by definition (de Vries et al. 2005).

A number of authors use SF definitions that are obviously some measure of AGN variability, but they are invalid in light of measuring the underlying true SF and hence ACF. The common problem is what kind of noise one should subtract when fitting an SPL SF. The rule of thumb should be as follows: for a $(m_i - m_j)^2$ term one subtracts $2\sigma_{\mathrm{n}}^2$ from it, and for a $(m_i - \langle m \rangle)^2$ term one subtracts just a single σ_{n}^2. Graham et al. (2014), apart from introducing a new method of variability analysis with the Slepian wavelets variance, combines a number of SF calculation procedures found in the literature. We have tested these SF definitions on the simulated data (see Figure 4).

$$\mathrm{SF}(\Delta t) = \langle (m_i - m_j)^2 \rangle \tag{14}$$

from Simonetti et al. (1984), Hook et al. (1994), and Choi et al. (2014) is in fact twice the definition of the "theoretical" SF from Equation (4). It does not subtract the noise term, and the square root of this equation (as in de Vries et al. 2005) is equivalent to Equation (10). Fitting an SPL SF to it gives flat SF slopes and cannot be used to infer the ACF. One would need to fit a full, four-parameter SF to obtain the correct result.

$$\mathrm{SF}(\Delta t) = \sqrt{\langle (m_i - m_j)^2 \rangle - \langle \sigma^2 \rangle} \tag{15}$$

from Bauer et al. (2009) subtracts an incomplete noise term (it should have a 2 in front of the noise term) and leads to flatter SFs (see Figure 4).

$$\mathrm{SF}(\Delta t) = \sqrt{\frac{\pi}{2} \langle |m_i - m_j| \rangle^2 - \langle \sigma^2 \rangle} \tag{16}$$

from di Clemente et al. (1996), Vanden Berk et al. (2004), and Bauer et al. (2009) subtracts an incomplete noise term (it should have a 2 in front of the noise term) and leads to flatter SFs.

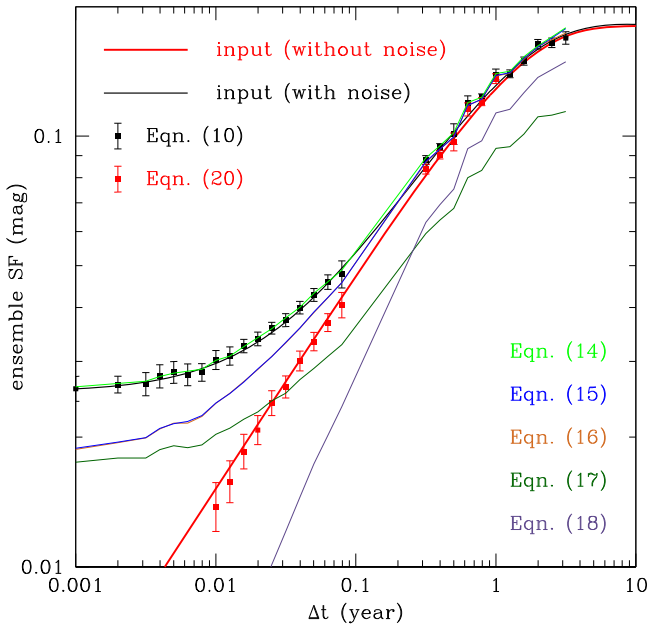

Figure 4. Example SF calculations for 1000 simulated AGN light curves in a bin $z = 1.6 \pm 0.1$ with $M_i = -26.0 \pm 0.25$ mag. In simulations we used the DRW model. The input signal (signal and noise) SF is shown as the thick red (black) line. The calculated SFs from the two methods used in this study are given as red and black points. We also show what common SF measures return (note that they should return the red line).

$$\mathrm{SF}(\Delta t) = \mathrm{median}[(m_i - m_j)^2] \tag{17}$$

from Sumi et al. (2005) does not subtract the noise term and returns flat SFs. In fact, the square root of it must be taken, and then it returns an SF shape similar to that from Equation (10), but the whole SF is shifted toward lower values.

$$\mathrm{SF}(\Delta t) = \left\langle \sqrt{\frac{\pi}{2}} |m_i - m_j| - \sqrt{\sigma_i^2 + \sigma_j^2} \right\rangle_{\Delta t} \tag{18}$$

from Schmidt et al. (2010a) is a strange measure; it generally looks as if it was subtracting too much of the noise term, leading to steeper and overall lower SF values. A plethora of SF definitions cause problems in their interpretations or comparisons; for example, both Schmidt et al. (2010b) and MacLeod et al. (2014) issue errata on SF definitions. Emmanoulopoulos et al. (2010) provide an overview of an SF analysis and in particular point out problems (unexpected breaks or wiggles) and caveats in its interpretation.

2.4. SF Calculations and Fitting

In this paper, we will be using two methods to estimate and model SFs:

(1) In the first one, we will use Equation (10), that is, without the photometric noise subtraction, to calculate the observed SF. We will fit to it the full, four-parameter SF described by the power exponential (PE) covariance matrix of the form $\sigma_{\mathrm{s}}^2 \exp(-(|\Delta t|/\tau)^{\beta})$. The four parameters include the power β, decorrelation timescale τ, variance at long timescales SF_{∞},

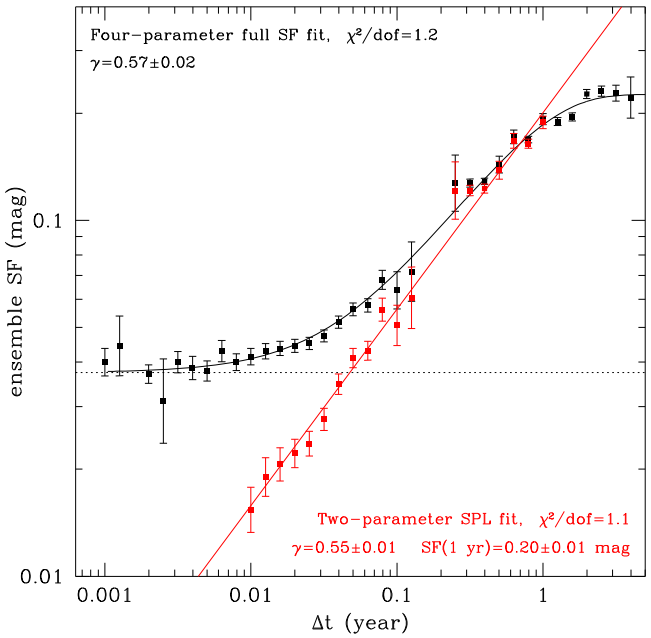

Figure 5. Example SF fits for 216 true SDSS AGNs in a bin $z = 1.6 \pm 0.1$ with $M_i = -26.0 \pm 0.25$ mag. The black and red points are calculated from Equations (10) and (20), respectively. The full, four-parameter SF fit (Equation (19)) is marked in black, while the SPL fit (Equation (13)) is marked in red. The dotted horizontal line marks the noise.

and the noise term σ_n:

$$\mathrm{SF}^2_{\mathrm{obs}}(\Delta t) = \mathrm{SF}^2_{\infty}(1 - \exp(-(|\Delta t|/\tau)^{\beta}) + 2\sigma^2_n. \quad (19)$$

We will also test a modification of the above fitting by fixing $\beta = 1$ (hence the three-parameter fit) and study the remaining SF parameters.

(2) We will also try another route where we calculate the distribution of Δm for $\Delta t < 2$ days and use it as the photometric noise estimate (because $\mathrm{SF}_{\mathrm{true}} \to 0$ as $\Delta t \to 0$), and we then subtract it in quadrature from Equation (10):

$$\mathrm{SF}^2_{\mathrm{true}}(\Delta t) = 0.549(\mathrm{IQR}^2(\Delta t) - \mathrm{IQR}^2(n)). \quad (20)$$

Subsequently, we will fit an SPL SF (Equation (13)) for lags in the range $4 < \Delta t < 365$ days, before the SFs start to flatten. Example SF calculations and fits of these methods on the real SDSS data are shown in Figure 5.

2.5. Power Spectral Density

In this paper, we will not measure the PSD simply because SFs are much easier to calculate and do not suffer from operational problems like irregular sampling and data binning. We briefly review the topic because in many AGN variability studies PSDs are the primary variability measure. We provide a means for their comparison to SFs and their connection to ACFs.

PSD(ν), a function of frequency ν, by definition is

$$\text{total power} \propto \int_{-\infty}^{\infty} \mathrm{PSD}(\nu) d\nu, \quad (21)$$

and PSD (ν) needs to be in units of power/Hz if the total power is to be expressed in units of power. Any time series $y(t)$, including AGN variability, can be thought of as a combination of separate periodic signals with frequencies ν. Using the Fourier transform, we can decompose such a time series into a series in the frequency domain using (e.g., Scargle 1981, 1982; Press et al. 1992c)

$$Y(\nu) = \int_{-\infty}^{+\infty} y(t) e^{-2\pi i \nu t} dt. \quad (22)$$

Parseval's theorem states that the total power in the time domain is equal to the total power in the frequency domain:

$$\text{total power} = \int_{-\infty}^{\infty} |y(t)|^2 dt = \int_{-\infty}^{\infty} |Y(\nu)|^2 d\nu. \quad (23)$$

The combination of Equations (21) and (23) gives

$$\mathrm{PSD}(\nu) = |Y(\nu)|^2, \quad (24)$$

which is simply the square of the Fourier transform of the signal and is expressed in units of $\mathrm{mag}^2\,\mathrm{Hz}^{-1}$.

The spectra of aperiodic signals can be represented as power laws, PSD (ν) $\propto \nu^{\alpha}$, where for example $\alpha = 0$ is the white noise (i.e., a constant PSD across all ν values). The amount of power in any two dex are the same for the pink or flicker noise ($\alpha = -1$), and the random walk noise (also Brownian noise) is described by $\alpha = -2$. A typical AGN PSD is shown in Figure 1.

A relation of the PSD to ACF can be found via the Fourier transform from the Wiener–Khinchin theorem that states

$$\mathrm{PSD}(\nu) = \int_{-\infty}^{\infty} \mathrm{ACF}(\Delta t) e^{-2\pi i \nu \Delta t} \Delta t. \quad (25)$$

Because the PSD measures the squared rms ($\mathrm{mag}^2\,\mathrm{Hz}^{-1}$) at a frequency ν, it can be thought of as "a reflection" of an SF (measuring plain rms), where the time lag Δt is replaced with a frequency ν and the conversion between powers is $\alpha = -4\gamma$. Please note that some authors present PSD $\times$ Hz instead of just PSD and that the conversion between powers $\alpha = -2\gamma - 1$ from Bauer et al. (2009) is for PSD $\times$ Hz and the square of SF (the "theoretical" SF; Equation (4)).

2.6. The DRW Model

Although it was noticed over half a century earlier (Ozernoi & Chertoprud 1966), Kelly et al. (2009) realized that the stochastic AGN variability, in particular the "red noise" (PSD $\propto \nu^{-2}$) at short lags and the white noise at long ones, has the same properties as the stochastic process called the DRW, also known as a first-order continuous-time autoregressive process (CAR(1)) or Ornstein–Uhlenbeck (OU; Uhlenbeck & Ornstein 1930) process. Subsequently, it has been shown that aperiodic optical light curves of individual quasars from ground-based studies can be well modeled by DRW (Kelly et al. 2009; Kozłowski et al. 2010a; MacLeod et al. 2010, 2011, 2012; Butler & Bloom 2011; Zu et al. 2011, 2013a; Ruan et al. 2012), while, at the "very short timescales" probed by the space-based *Kepler* mission and apparently weakly probed by ground-based studies, the PSD shows a departure from DRW (Mushotzky et al. 2011; Kasliwal et al. 2015a; we will discuss this in Section 4). Kelly et al. (2014) provide a more general approach to modeling unevenly sampled stochastic light curves using the CARMA models. DRW is the simplest of the CARMA(p, q) models, with $p = 1$ and $q = 0$, and as a matter of fact only DRW can be identified with the physical processes such as Brownian motions or the Wiener process having a natural interpretation of the model parameters.

A quasar light curve can be described (and fitted) with the DRW model, having just two model parameters: the damping

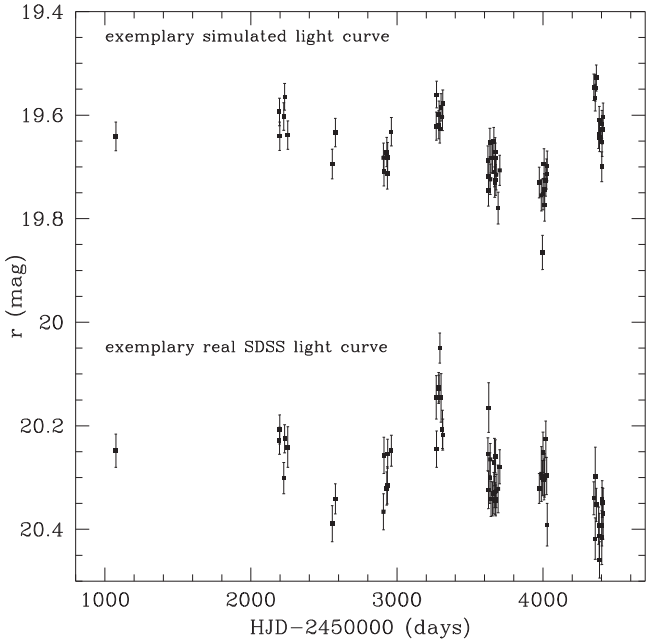

Figure 6. Example light curves for a simulated (top) and true SDSS (bottom) AGN variability.

timescale τ (the same decorrelation timescale as in SFs) and the modified variability amplitude $\hat{\sigma} = \sigma\sqrt{2/\tau}$ (or equivalently $\mathrm{SF}_\infty = \hat{\sigma}\sqrt{\tau} = \sqrt{2}\,\sigma$; MacLeod et al. 2010). The ACF for DRW has the form

$$\mathrm{ACF}(\Delta t) = e^{-|\Delta t|/\tau} \tag{26}$$

and can be generalized as the PE (e.g., Zu et al. 2013a):

$$\mathrm{ACF}(\Delta t) = e^{-(|\Delta t|/\tau)^{\beta}}, \tag{27}$$

for $0 < \beta < 2$ and $\tau > 0$, where $\beta = 1$ corresponds to DRW (Figure 2). The parameter β is responsible not only for the shape of the ACF but also for the SF and PSD slopes. Small values of β drive the correlation to be weaker, which in turn produces flatter SFs (Figure 2). In opposition, large values of β make the correlation stronger, which causes the SF to steepen.

In order to understand both the SFs and the recovered DRW parameters from the real SDSS data, we will perform simulations of light curves using DRW. We generate 100-year-long light curves (to make sure the signal covariance between points is zero, also known as the red noise leakage) using the prescription of Kozłowski et al. (2010a). The light curves will be later cut into shorter time baselines or have the cadence reduced or include seasonal gaps, nominally degraded to a typical ground-based survey (Figure 6). In short, the chain is initiated by $s_1 = G(\sigma^2)$, where $G(x^2)$ is a Gaussian deviate of dispersion x. The subsequent light curve points come from

$$s_{i+1} = s_i e^{-\Delta t/\tau} + G[\sigma^2(1 - e^{-2\Delta t/\tau})], \tag{28}$$

where $\Delta t = t_{i+1} - t_i$ is the time interval. The observed light curve is obtained from $y_i = s_i + G(n_i^2)$, where n_i is the observational noise. Simulations of generic light curves for any covariance functions are explained in, e.g., Zu et al. (2011) and Kozłowski (2016).

3. REVISITING AGN VARIABILITY IN SDSS

We will now analyze a sample of ∼9000 SDSS AGNs, considered by MacLeod et al. (2010), to study their variability through the SF analysis and DRW light curve modeling. Of the five SDSS filters, the *r*-band has the highest throughput (Fukugita et al. 1996), and we will be primarily considering this filter. The light curves come from data reductions by Ivezić et al. (2007). The AGN identifications are taken from the SDSS Data Release 5/7 (Schneider et al. 2007; Abazajian et al. 2009), while the basic physical parameters such as the absolute magnitudes come from Schneider et al. (2007), and the black hole masses and bolometric luminosities are from Shen et al. (2008). Whenever necessary, we calculate our own K-correction using the SDSS filter throughputs and the mean AGN spectrum from Vanden Berk et al. (2001). We have used the standard cosmological LCDM model with $(H_0, \Omega_M, \Omega_{\rm vac}, \Omega_k) = (70\ \mathrm{km\ s^{-1}\ Mpc^{-1}}, 0.3, 0.7, 0.0)$.

3.1. Data Analysis

To subtract the noise term correctly (or to fit it correctly) or for it to have well-behaved properties, we will be considering a "subensemble" variability of AGNs in the narrow redshift–absolute magnitude bins of size 0.2–0.5 mag, with more than 10 AGNs per bin (median 50 AGNs). Emmanoulopoulos et al. (2010) provide an overview of an SF analysis and notes that breaks and wiggles at long timescales in a single SF can occur even in a process not having such an intrinsic break timescale. (We observe an analogous behavior, but we also find that averaging a large number of SFs causes such breaks and wiggles to vanish, which in a statistical sense means that they are no longer outliers from the input SF (see Figure 4).)

Each AGN light curve is converted into a file with the time differences between epochs and the corresponding differences of magnitudes for these epochs (dt_i, dm_i), where the time differences are corrected for redshift. In each redshift–absolute magnitude bin, we combine all AGN files and calculate distributions of Δm in logarithmic bins of the time difference Δt of width 0.1 dex and use Equation (10) to estimate the rms variability at each Δt. Please note this is now the SF that includes the noise term.

We estimate the SF error bars in each Δt bin by simulation means. (1) We have tested the bootstrap method. (2) We have tested the dispersions of IQRs of randomly selected subsamples of the original distribution. (3) We generated a large number of pure Gaussian distributions with N data points and empirically found that the uncertainty of the SF measurement in a bin, calculated as $\mathrm{SF} = 0.741 \times \mathrm{IQR}$, is simply $\Delta\mathrm{SF} = 1.17 \times \mathrm{SF} \times N^{-1/2}$ and equals the dispersion of the SF for the various Gaussian distribution realizations. Because in every Δt bin we combine the magnitude differences from a number of different AGN light curve realizations from a narrow but nonzero spread of magnitudes, we find that the bin-to-bin SF variations from the three methods are much larger than these estimated uncertainties. The optimal noise estimate is different, and we adopt $\Delta\mathrm{SF} = 4 \times N^{-1/2}$ as our SF uncertainty in order to obtain $\chi^2/\mathrm{dof} \approx 1$ across the full redshift–absolute magnitude plane for the SF fits to the simulated data. Most importantly, such an approach returns correctly measured SF parameters (see the simulations in Figures 7–8).

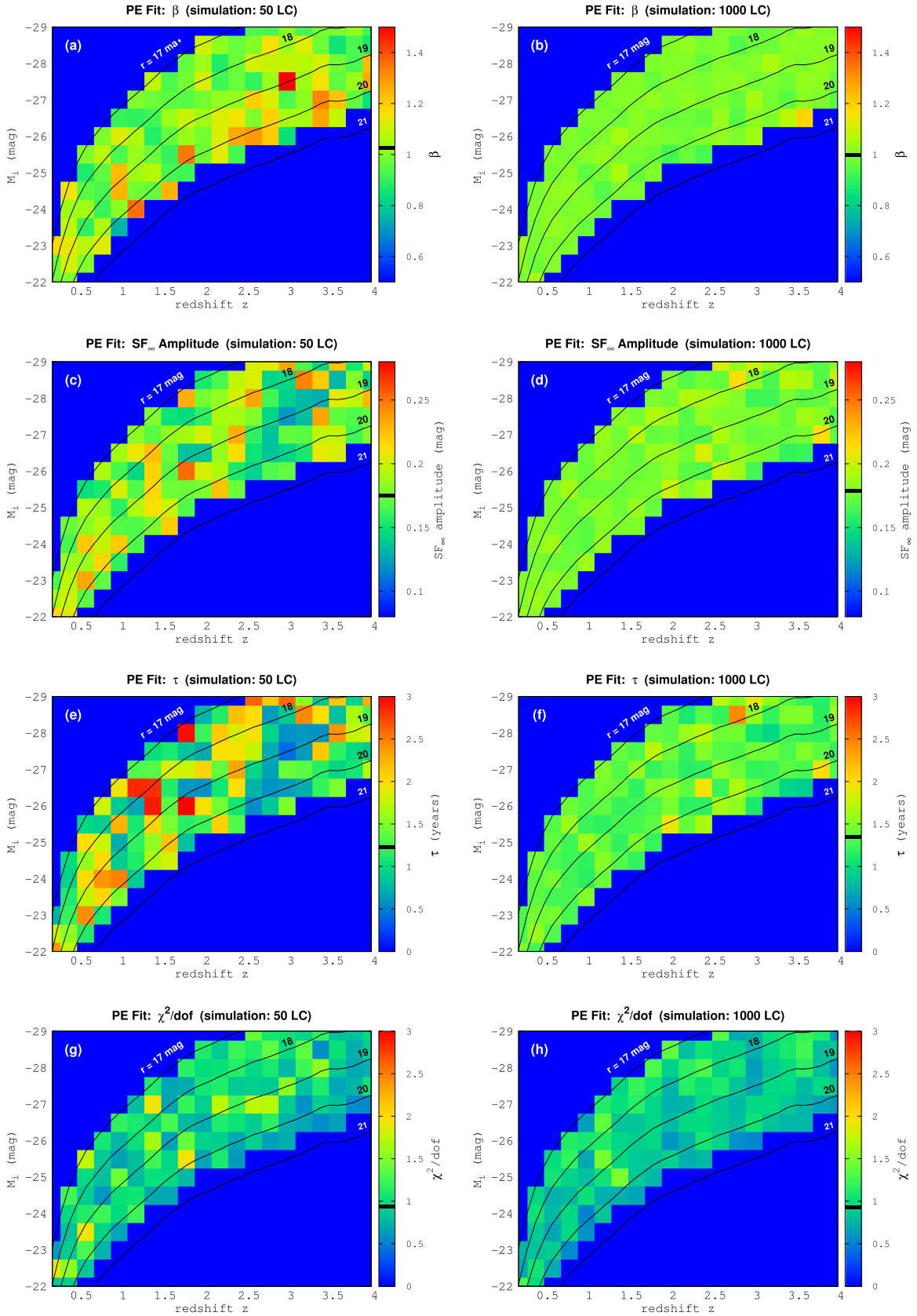

Figure 7. The recovered SF parameters from simulations of 50 (left column) and 1000 (right) AGN light curves per redshift–absolute magnitude bin. The light curves were simulated as the DRW stochastic process ($\beta = 1.0$) with the input parameters $\tau = 500$ days (1.37 years) and $\mathrm{SF}_\infty = 0.18$ mag. All panels show the same redshift–absolute magnitude ranges and also lines (black) for the constant observed *r*-band magnitude. Each panel is complemented by a dedicated color scale on its right that spans an appropriate range of the parameter space. The median values (weighted with the AGN number) are marked on the color bars with the thick black line. The light curves were fit with the full, four-parameter SF function that measures the power β of the power exponential (PE) covariance matrix of the signal. The recovered parameters are nearly identical to the input ones and hence are unbiased functions of the absolute magnitude and redshift.

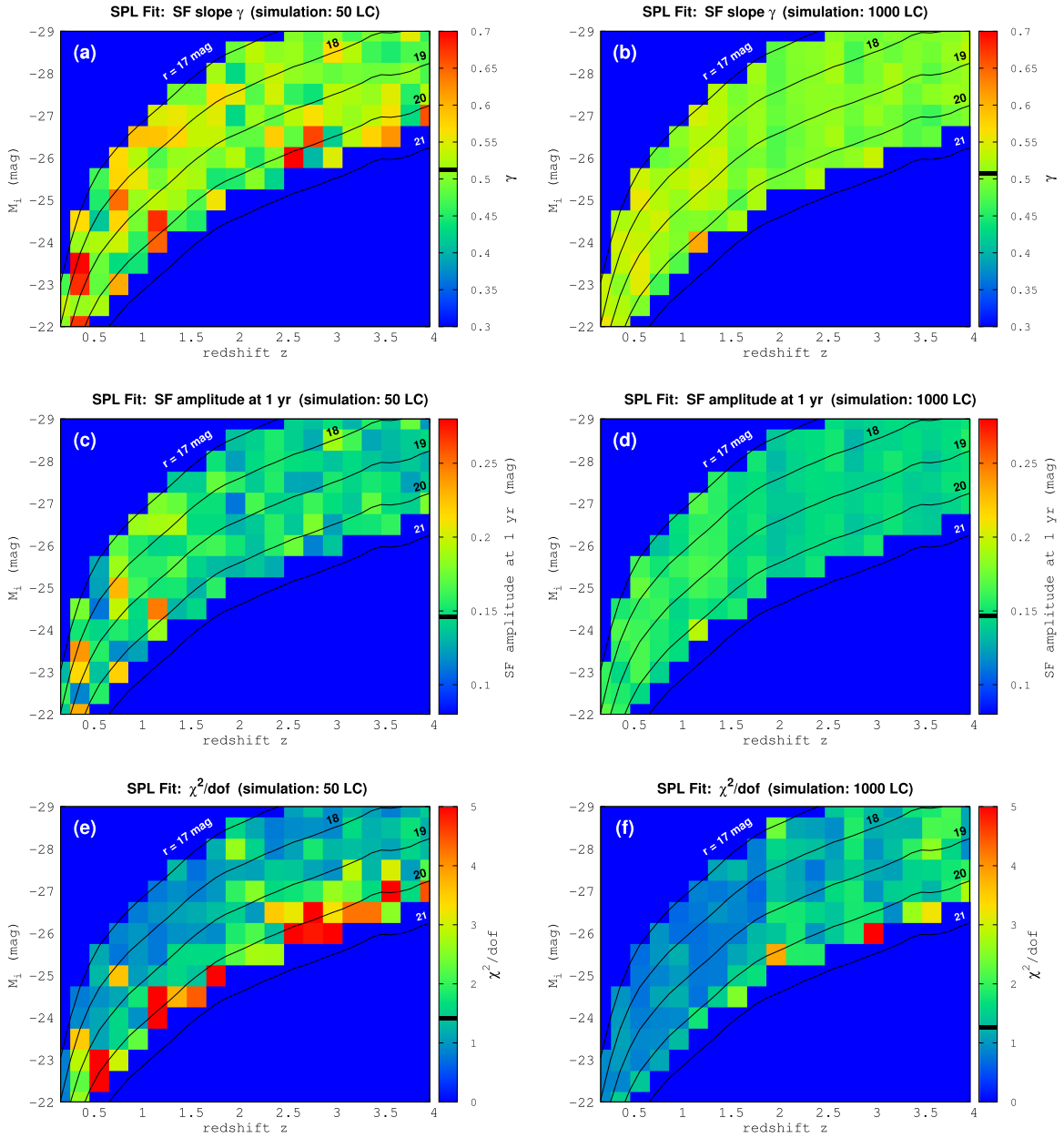

Figure 8. The recovered SF parameters from simulations of 50 (left column) and 1000 (right) AGN light curves per redshift–absolute magnitude bin. The light curves were simulated as the DRW stochastic process ($\gamma = 0.5$) with the input parameters $\tau = 500$ days and $SF_\infty = 0.18$ mag (SF at one year is 0.15 mag). All panels show the same redshift–absolute magnitude ranges and also the lines (black) of the constant observed *r*-band magnitude. Each panel is complemented by a dedicated color scale on its right that spans an appropriate range of the parameter space. The median values (weighted with the AGN number) are marked on the color bars with the thick black line. The light curves were fit with a single power law (SPL) SF function in the range $4 < \Delta t < 365$ days. There is a small bias in the recovered SPL slope $\gamma = 0.52$, as compared to the input 0.5, that slightly increases toward lower redshifts. The amplitude shows a similar behavior, but it is smaller at higher redshifts than the input value.

We will be considering two fitting methods now (described in Section 2.4):

(1) We will fit the SF using the full, four-parameter SF model with Equation (19) and employ the MINUIT minimization routines.[1] The SDSS S82 light curves are generally short (in the rest frame), weakly probing the SF turnover, so in the minimization procedure we have added a weak prior on the timescale, $\Delta\chi^2 = (\tau - 1.5\ \mathrm{year})^2$, to prevent it from going to infinity, where ∼1.5 year is the typical AGN timescale found from DRW modeling in MacLeod et al. (2010). In the majority of the bins the timescale is recovered correctly without this additional prior.

Next, we make a small modification to method (1), where we fix $\beta = 1$. It is clear we will not measure the SF slope, but we are going to study the remaining SF parameters. In particular, from method (1) (as we will show later) we see that the SF slope slightly steepens with the increasing luminosity, but the SF decorrelation timescale stays constant. From direct light curve modeling with DRW, which has the fixed $\beta = 1$, we see an increase of the timescale with the increasing luminosity. This method will be primarily used to check if the DRW timescale increases with the increasing β (and the answer ahead is yes).

[1] www.cern.ch/minuit

(2) In this method, we calculate distributions of Δm for $\Delta t < 2$ days and use it as a noise estimate that we subtract in quadrature from the SFs in method (1); see Equation (20). This is a noise-free, true SF that can be fitted with an SPL, Equation (13), in the timescale range $4 < \Delta t < 365$ days, that is, before the SF starts to flatten in the vicinity of one year. Because this is a simple line fitting, we use a least-squares χ^2 minimization (as explained by, e.g., Gould 2003).

Example calculated SFs and best fits obtained with methods (1) and (2) are shown in Figures 4–5.

3.2. Simulations of the SDSS Data

Prior to estimating the variability parameters from the real data, we have to understand any possible biases or systematic effects present in the methods used to measure the SFs. We do this by simulating artificial AGN light curves in two samples with 50 and 1000 AGNs per bin. The first one will show us approximately the expected scatter between bins for the real data, while the latter will enable us to track down any low-level systematics and biases. We simulate 100-year-long light curves with a cadence of four days. From such a long time series, we pick only these epochs matching the real ones (typically 60 epochs) that occur at least 40 years after the simulation starts. This is to make sure all correlations at long Δt are included (and there is no red noise leakage). The exact procedure for generating the light curves is detailed in Section 2.6. They are simulated with the DRW model, so the expected SF slope for $\Delta t \ll \tau$ is $\gamma \equiv 0.5$. The input parameters are picked to be $\tau = 500$ days and $\mathrm{SF}_\infty = 0.18$ mag (in fact, $\tau = 500(1+z)$ days, and then the simulated light curves are corrected for the $(1+z)$ term).

The results of the simulations are presented in Figures 7–8, where the left (right) column presents the subensemble variability results for 50 (1000) AGN light curves per bin. We will be interested in any biases or systematics between the measured output variability parameters and the input ones that are known. In Figures 7–8, each panel is complemented by a dedicated color bar spanning an adequate parameter range. To estimate the median value across the absolute magnitude–redshift plane, we create a distribution of values by counting the measured parameter in every bin a number of times equal to the number of AGNs in that given bin. The median value from such a distribution is measured and marked as a thick black line in the color bars of Figures 7–8.

In Figure 7, we see that, for the simulations with both 50 and a 1000 light curves per bin, the recovered parameters from the full SF fitting are stable and unbiased functions of redshift and the absolute magnitude. The median $\beta = 1.016$ ($\gamma = 0.5008$) is nearly identical to the input value $\beta_{\mathrm{input}} = 1.0$, with median $\mathrm{SF}_\infty = 0.175$ mag and 0.179 mag for the 50 and 1000 light curves, respectively, while $\mathrm{SF}_{\infty\ \mathrm{input}} = 0.18$ mag. The input timescale is $\tau_{\mathrm{input}} = 500$ days (1.37 year), and the measured values are $\tau = 1.22$ and $\tau = 1.35$ years for the 50 and 1000 light curves.

In Figure 8, we present the best-fit parameters obtained from an SPL SF fitting. We observe very weak trends for γ and the SF amplitude at one year with redshift. The distribution of recovered γ peaks at 0.52 for bins with 50 light curves, slightly higher than the DRW input of $\beta_{\mathrm{input}} = 1.0$. We will be later correcting the γ obtained from the real data for this small bias.

3.3. The Real SDSS Data

In panels (a) and (b) of Figures 9 and 10, we present the mean values (in a bin) of the black hole mass and the Eddington ratio, respectively. It is clear that both these parameters increase with increasing AGN luminosity. In panels (c) and (d) of Figure 9, we present the measured values of the power β of the PE covariance matrix and the amplitude at long timescales ($\Delta t \to \infty$), SF_∞, respectively. It appears that β slightly increases (SF_∞ decreases) with increasing luminosity. From the simulations, we know these parameters should be unbiased functions of redshift and luminosity. The measured decorrelation timescales seem to be constant across the redshift–absolute magnitude plane, panel (e) of Figure 9, and the goodness of fits seems to be also reasonable, panel (f). We also model SFs with a three-parameter fit (with fixed $\beta = 1$). In panels (g) and (h), we show the recovered timescales with and without the weak prior on the timescale ($\Delta\chi^2 = (\tau - 1.5\ \mathrm{yr})^2$), respectively. We see that by fixing $\beta = 1$, the steepening of β in the four-parameter fit is now replaced by the increasing timescale, in a similar fashion to what is observed for direct light curve modeling with DRW (Figure 10, panel (g)).

In panels (c) and (d) of Figure 10, we present the SF slope γ and the amplitude at one year, measured from the SPL fitting in the timescale range $4 < \Delta t < 365$ days. We see a similar dependence of γ with the luminosity as in the case of β in panel (c) of Figure 9; that is, γ increases when the luminosity increases. Also similarly to SF_∞ in panel (d) of Figure 9, the amplitude at one year decreases with the increasing luminosity in panel (d) of Figure 10. Panels (e) and (f) present the number of SDSS Stripe 82 AGNs per redshift–absolute magnitude bin and the goodness of fits, respectively. We also modeled the SDSS light curves directly using the DRW model (see Kozłowski et al. 2010a for details). In panels (g) and (h), we present the mean values of the DRW parameters τ and $\hat{\sigma}$, respectively. It is clear that the DRW timescale τ shows a similar dependence on redshift and absolute magnitude as parameters β and γ, while the $\hat{\sigma}$ is akin to the amplitude at one year and SF_∞.

We also perform simulations of 1000 DRW light curves per bin to measure the biases in the parameters τ and $\hat{\sigma}$ in the redshift–absolute magnitude plane (Figure 11). From panel (a), it is very clear that the recovered timescales are typically underestimated as compared to the input ones, and this ratio decreases with the increasing ratio of the input timescale to the experiment length. In panel (b), we show the parameter $\hat{\sigma}$. It is typically only weakly biased toward larger values (by 3%) and rises significantly when the photometric noise becomes comparable to the variability amplitude (for faint objects).

4. DISCUSSION

We identified a number of SF definitions in the literature that seem to be flawed, so the reported correlations of variability with the physical parameters that are based on them may also need revision. In particular, is the amplitude really correlated with the luminosity, rest-frame wavelength, or redshift, or is it simply a correlation with the (partially subtracted) photometric noise? We inspect some of the basic correlations below.

Because in simulations we observe no correlations of the SF parameters with redshift or the absolute magnitude (Figure 7

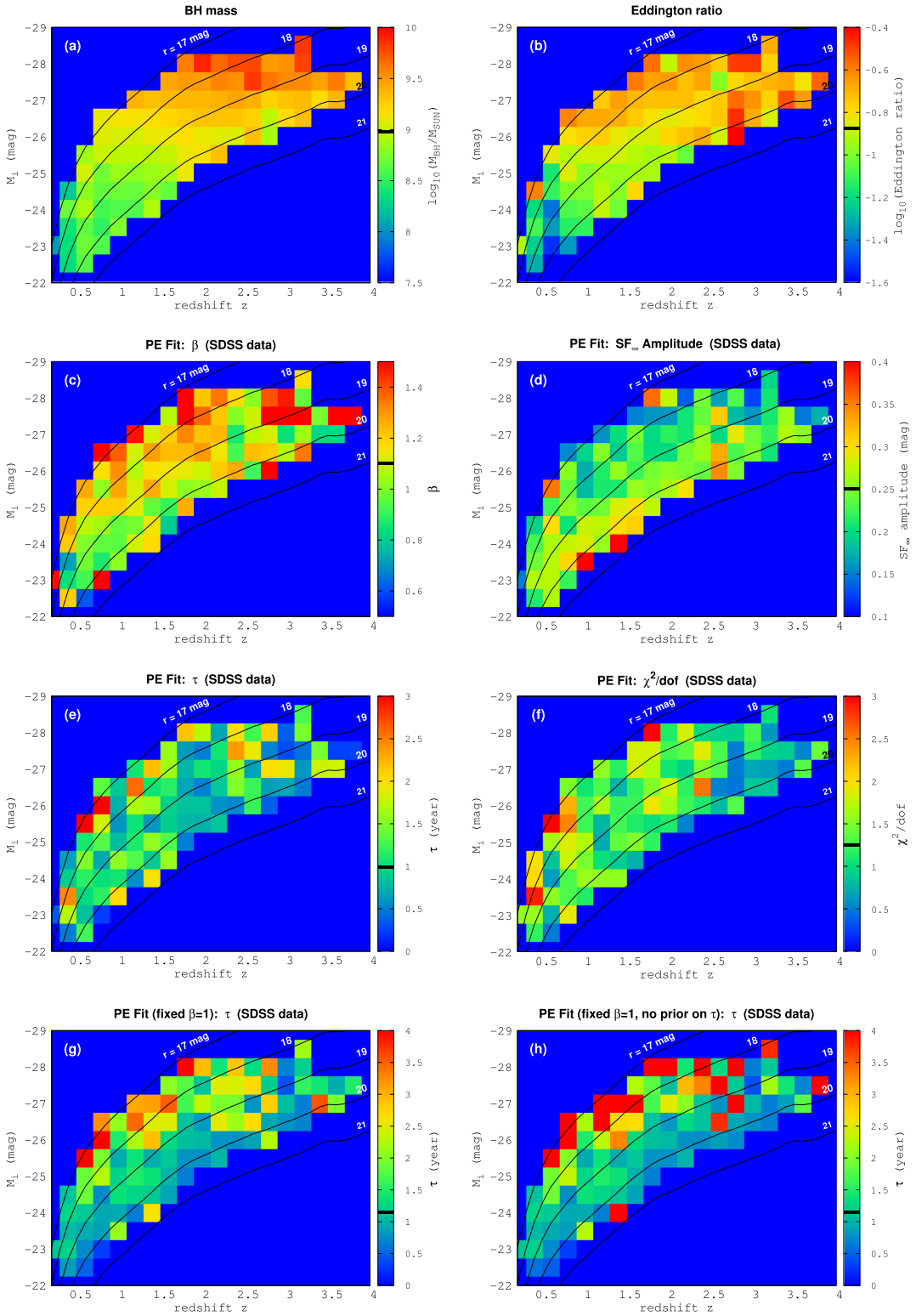

Figure 9. Variability analysis of ∼9000 SDSS Stripe82 AGNs in the *r*-band. All panels show the same redshift–absolute magnitude ranges and also the lines (black) of the constant observed *r*-band magnitude. In panels (a) and (b), we present the mean black hole mass and the mean Eddington ratio, respectively. We fit the full, four-parameter SF, where we model the ACF as the power exponential. The model is given by Equation (19) and has four parameters: the shape of the ACF β (panel (c)), the variability amplitude at long time lags SF_∞ (panel (d)), the timescale of decorrelation τ (panel (e)), and the noise term (not shown). The goodness of fit is presented in panel (f). We also present the timescale τ for the three-parameter fit (where the fourth parameter is fixed $\beta = 1$) in panels (g) and (h). Nominally, we use a weak prior on the timescale in the three- and four-parameter fits to avoid infinities; in panel (g) we present the timescale with the prior and in panel (h) without the prior. The χ^2/dof for these fits slightly rises with increasing τ from ∼1.5 to 2.0–2.5 (not shown). Each panel is complemented by a dedicated color scale on its right that spans an appropriate range of the parameter space. The median values (weighted with the AGN number) are marked on the color bars with the thick black line.

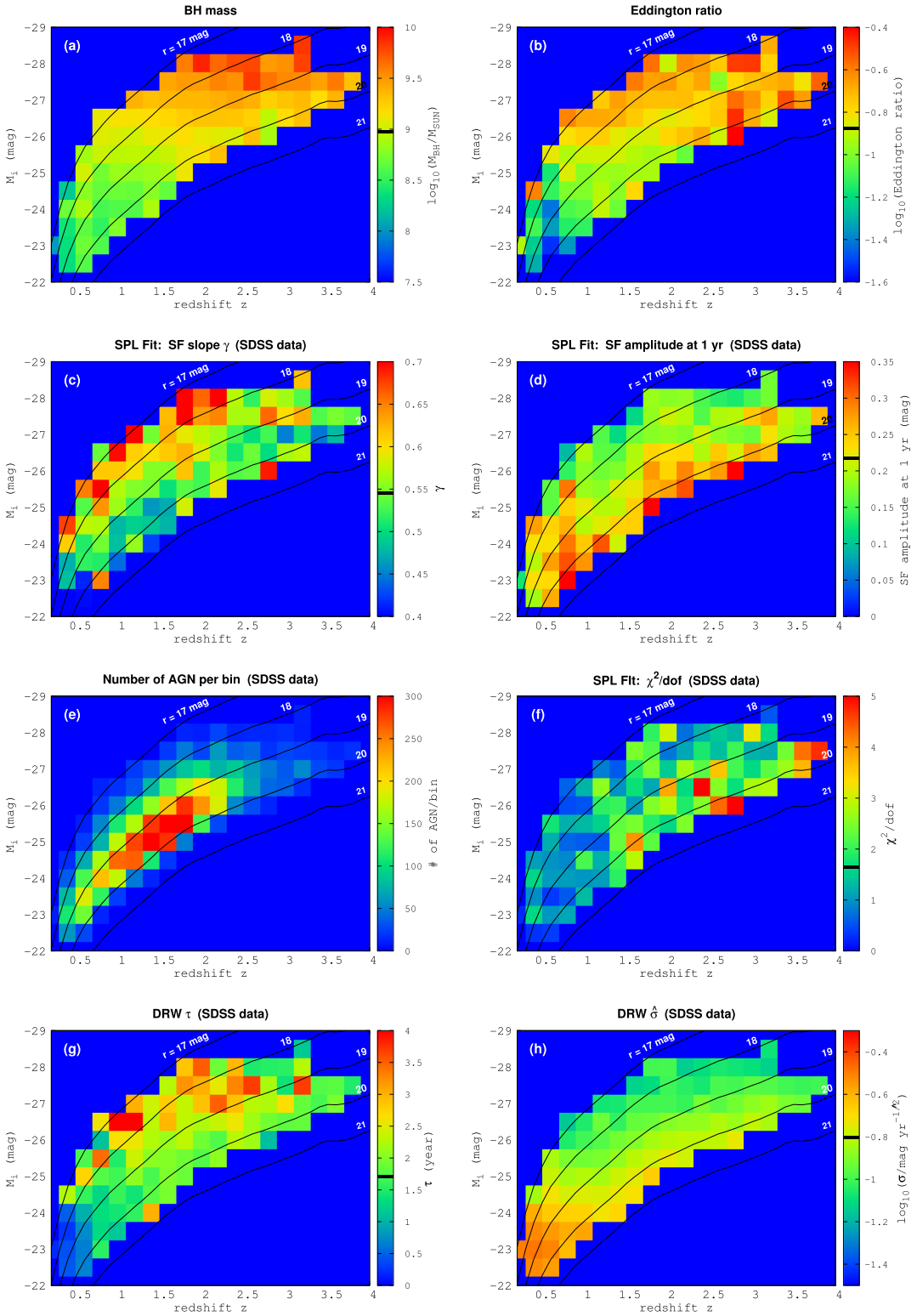

Figure 10. Variability analysis of ∼9000 SDSS Stripe82 AGNs in the *r*-band. All panels show the same redshift–absolute magnitude ranges and also the lines (black) of the constant observed *r*-band magnitude. In panels (a) and (b), we present the mean black hole mass and the mean Eddington ratio, respectively. We fit the two-parameter single power-law SF (Equation (20)) in a range $4 < \Delta t < 365$ days and in panels (c) and (d) show the resulting parameters γ and the amplitude at one year, respectively. In panel (e), we show the number of AGNs per bin, while in panel (f) we present the χ^2/dof for the SF fit. We also model each light curve with DRW and present the mean timescale τ and the mean modified amplitude $\hat{\sigma}$, shown in panels (g) and (h), respectively. They are correlated with the SF γ and the amplitude at one year, respectively. Each panel is complemented by a dedicated color scale on its right that spans an appropriate range of the parameter space. The median values (weighted with the AGN number) are marked on the color bars with the thick black line.

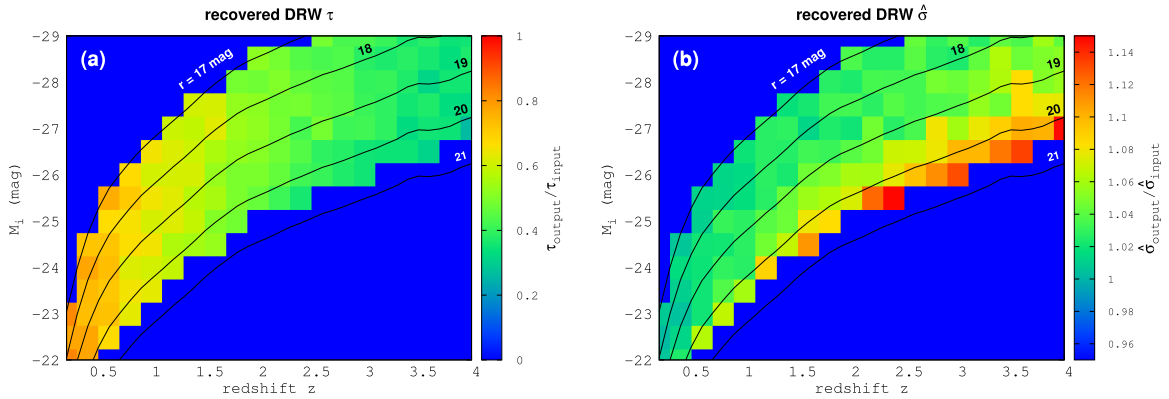

Figure 11. DRW simulations (with $\tau = 500$ days and $SF_\infty = 0.18$ mag) and modeling of 1000 light curves per bin. Ratios of the output to input parameters are shown. The returned timescales (panel (a)) are increasingly underestimated with increasing redshift (in fact with the increasing ratio of the timescale to the experiment length), and there is only a weak overestimate (~3%) of $\hat{\sigma}$ for $r < 19.5$ mag (panel (b)), slightly increasing with redshift. The longer the experiment span, the more reliable the estimate of the true DRW timescales. However, any recovered DRW parameters should be debiased by simulation means.

and 8), the real data (Figure 9 and 10) should provide unbiased measures of these parameters.

4.1. Correlations

To untangle dependencies of the variability parameters on the physical parameters, we will consider two cases: (1) a subsample of AGNs with fixed luminosity and (2) a subsample of AGNs with fixed black hole mass.

(1) We selected a sample of 518 AGNs with $19.5 < r < 20.0$ mag (narrow luminosity range) and $1.3 < z < 1.7$ (similar redshift/emission wavelength) and divided them into five bins in $\eta_{\rm Edd}$, each containing over a hundred AGNs. We fitted the subensemble SFs with the full SF fit (method (1)) and found a lack of correlation of β with the Eddington ratio or the black hole mass (Table 1), but we found an anticorrelation of the asymptotic amplitude with the Eddington ratio of the form $\log({\rm SF}_\infty) \propto (-0.28 \pm 0.06)\log(\eta_{\rm Edd})$, in agreement with the value found by MacLeod et al. (2010), that is, -0.23 ± 0.03, and in rough agreement with Wilhite et al. (2008), who measured ~ -0.15. Please note, however, that they used the incomplete SF equation from di Clemente et al. (1996).

Because we are inspecting here a narrow luminosity range and $-1.5 < \log(\eta_{\rm Edd}) < 0$, the change in the Eddington ratio ($\eta_{\rm Edd} \propto L_{\rm bol} M_{\rm BH}^{-1}$) is due to the changing black hole mass in a range $8 < \log(M_{\rm BH}/M_\odot) < 9.5$. We observe a correlation of the amplitude with the black hole mass of the form $\log({\rm SF}_\infty) \propto (0.29 \pm 0.04)\log(M_{\rm BH})$, slightly higher than that (0.18) found in MacLeod et al. (2010).

The timescale τ is correlated with the black hole mass $\log(\tau) \propto (0.38 \pm 0.15)\log(M_{\rm BH})$, with a somewhat higher index (0.21) than in MacLeod et al. (2010), and is anticorrelated with the Eddington ratio $\log({\rm SF}_\infty) \propto (-0.36 \pm 0.15)\log(\eta_{\rm Edd})$ (Table 1).

(2) We selected 837 AGNs in a range $8.7 < \log(M_{\rm BH}/M_\odot) < 9.3$ (constant black hole mass) and $1.3 < z < 1.7$ (similar redshift/emission wavelength) and divided them into six luminosity classes with each containing 50–200 AGNs. We fitted the subensemble SFs with the full SF fit (method (1)). We observe a slight increase of β with the increasing bolometric luminosity (an SPL index 0.10) and the Eddington ratio (0.16). The asymptotic variability $\log({\rm SF}_\infty) \propto (-0.35 \pm 0.05)\ \log(L_{\rm bol})$ and $\log({\rm SF}_\infty) \propto (-0.55 \pm 0.06)\ \log(\eta_{\rm Edd})$ (Table 1). The timescale is nearly independent of the bolometric luminosity and the Eddington ratio, $\log(\tau) \propto (-0.05 \pm 0.17)\ \log(L_{\rm bol})$ and $\tau \propto (-0.16 \pm 0.25)\ \log(\eta_{\rm Edd})$, respectively, and is consistent with being constant as reported by MacLeod et al. (2010), while staying in contrast to the positive correlation in Hawkins (1993).

Table 1
Correlations of the AGN Variability and Physical Parameters

Variability Parameter	Physical Parameter	Power-law Index
	Constant Luminosity	
$\log(\beta)$	$\log(M_{\rm BH})$	0.01 ± 0.05
$\log(\beta)$	$\log(\eta_{\rm Edd})$	-0.01 ± 0.04
$\log({\rm SF}_\infty)$	$\log(M_{\rm BH})$	0.29 ± 0.04
$\log({\rm SF}_\infty)$	$\log(\eta_{\rm Edd})$	-0.28 ± 0.06
$\log(\tau)$	$\log(M_{\rm BH})$	0.38 ± 0.15
$\log(\tau)$	$\log(\eta_{\rm Edd})$	-0.36 ± 0.15
	Constant black hole mass	
$\log(\beta)$	$\log(L_{\rm bol})$	0.10 ± 0.03
$\log(\beta)$	$\log(\eta_{\rm Edd})$	0.16 ± 0.05
$\log({\rm SF}_\infty)$	$\log(L_{\rm bol})$	-0.35 ± 0.05
$\log({\rm SF}_\infty)$	$\log(\eta_{\rm Edd})$	-0.55 ± 0.11
$\log(\tau)$	$\log(L_{\rm bol})$	-0.05 ± 0.17
$\log(\tau)$	$\log(\eta_{\rm Edd})$	-0.16 ± 0.25

Note. For $\Delta t \ll \tau$ the slope of the SF γ is related to the power β as $\beta \equiv 2\gamma$, so relations for β can be directly translated to SF slopes.

4.2. Comments on DRW

A number of studies (e.g., MacLeod et al. 2012; Zu et al. 2013a; Kozłowski et al. 2016) based on *Kepler* results from Mushotzky et al. (2011) stated that DRW is not the right model for "very short timescales" that are not or are weakly probed by ground-based surveys. This is only partly true; namely, *Kepler* PSDs appear to be steeper, but the frequencies involved, 10^{-8}–10^{-5} Hz or 1.2–115 days, are in fact well probed by ground-based studies. In particular, the OGLE survey with the cadence of 2–4 days (Udalski et al. 2015) has been the gold standard for many astrophysical fields for over two decades, and Zu et al. (2013a) was unable to find deviations from DRW in such data. Note, however, that

Kozłowski (2016) recently showed that direct modeling of light curves with DRW is equally good for both DRW and non-DRW processes and cannot be used to identify what process is being modeled. There is rather a full discrepancy between *Kepler* and the ground-based surveys, than the claimed one on "very short timescales" (*Kepler* PSDs hit the white noise at (or below) the 1-day cadence, so it does not probe the covariance of the signal below these timescales). So Kasliwal et al. (2015b) justly asked if *Kepler*'s light curves need reprocessing, and they positively answered this question. In fact, steeper PSDs or SFs mean a faster variability that is "easy to achieve" by improper photometric procedures, such as when too much light/background is subtracted from a light curve.

Kelly et al. (2011) considered AGN variability in light of a linear combination of multiple OU processes. Such a combination of DRW processes may lead to an average ACF with powers different from $\beta = 1$, although individual processes are DRW ($\beta = 1$). So, for example, a shallower or steeper SF may mean either one process that is different from DRW or a combination of DRW processes leading to shallower or steeper SFs.

We will now discuss a DRW modeling of the simulated SDSS light curves with constant $\tau = 500$ days and $SF_\infty = 0.18$ mag, but for a range of redshifts and absolute magnitudes. For each absolute magnitude–redshift bin there are 1000 AGN light curves generated and modeled with DRW (Figure 11). Each model parameter is represented by a posterior probability distribution, where the peak is identified here with the best value and the width of the distribution (the parameter uncertainty) is not considered. Of the 1000 best parameter values per bin we calculate their median. From Figure 11, we see that the measured DRW parameters are a strong function of the ratio of the input timescale (which increases with redshift) to the total experiment span. What happens if you cut an AGN light curve, let us say, in half? Well, because the ratio of τ to the experiment length increases, the returned timescale decreases. So a longer light curve, with the same underlying process as the shorter one, will have a longer (and closer to the true value) timescale than the cut light curve. Because this was not accounted for, it could have been the reason for problems in finding time lags due to light time travel in "the photometric reverberation mapping" of the OGLE-III (eight years long) and OGLE-IV (then four years long) light curves in Zu et al. (2016).

4.3. Typical AGN Variability

In Figures 9 and 10, the SDSS S82 data were divided into bins primarily to avoid mixing AGNs with the different photometric noise, which causes problems in subtracting or modeling it, but also to investigate any possible correlations of variability with the physical parameters of AGNs. We will now combine the results from these figures to obtain "typical" AGN variability parameters.

Because bins contain different numbers of AGNs (panel (e) in Figure 10), imagine a simple situation where we consider only two bins, one with 10 AGNs with an "incorrect" SF value and the second bin with 300 AGNs and the "correct" SF value. If they had equal weights, then the average of the two would be off from the "correct value." Therefore we will add weights in averaging bins: when constructing histograms, we simply count each bin a number of times equal to the number of AGNs in it.

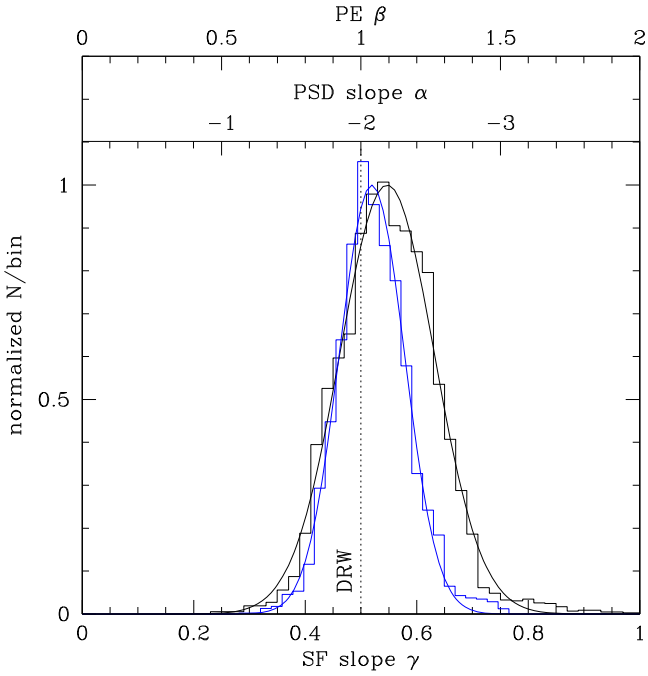

Figure 12. Histograms, weighted with the number of AGNs in a bin and normalized so the Gaussian fits peak at 1, of SF slopes γ (PE β and PSD slope α on top) from fitting the full SF function (black) and a single power-law SF for $4 < \Delta t < 365$ days (blue) to the redshift–absolute magnitude bins (corrected for biases). The histograms can be modeled relatively well as Gaussians with the mean values $\gamma = 0.55$ and 0.52, and dispersions $\sigma_\gamma = 0.08$ and 0.06, respectively. The DRW model ($\gamma = 0.5$) is marked with the vertical dotted line.

In Figure 12, we present weighted histograms of the SF slope γ. They were obtained from the two fitting methods, the full-fit SF slopes returning $\beta \equiv 2\gamma$, and the SPL SF fits. These histograms peak at $\gamma = 0.55 \pm 0.08$ and 0.52 ± 0.06, respectively. These values are a little higher than the expected $\gamma = 0.5$ from DRW but are entirely consistent with it. The timescale τ has a wide and non-Gaussian distribution. The median value is $\tau = 0.98$ year, and we estimate the error bar using the IQR; it is $\sigma_\tau = 0.41$ year. If we fitted this histogram with a Gaussian, it yields $\tau = 0.97 \pm 0.46$ year. The weighted histogram for the asymptotic amplitude is $SF_\infty = 0.25 \pm 0.06$ mag, while the one for the amplitude at one year is $SF_0 = 0.22 \pm 0.06$ mag. Note, however, that the SFs are already slightly flattening at this time lag, so the latter value is simply overestimated. The true value is $SF_0 = 0.20$ mag.

5. SUMMARY

In this paper, we have reviewed basic variability observables that are often used in optical and infrared AGN studies. From basic principles we show how to correctly measure the SF so it can be directly linked to the covariance function, or equivalently the autocorrelation function, of the underlying stochastic process (or processes) causing the variability (please refer to Kelly et al. 2009, 2011, 2014 for details on the stochastic processes).

Prior to the measurement of the AGN variability parameters from the SDSS S82 data, we simulate similar data sets where the input parameters are known. By comparison of the output and input parameters, we investigate biases and systematics that can potentially be present in the SF measurement from the

real data. Also by simulation means we review commonly used SF measures and point out their problems. The most common one is the subtraction of an incomplete or no noise term, leading to flat SF slopes. SFs can be successfully modeled as an SPL for the rest-frame time lags in the range between days and about one year (before turning into white noise), provided the noise term is subtracted carefully (and fully).

The key result of this paper is the slope of the SF measured from two methods: the full SF modeling using the PE covariance matrix and the SPL SF fitting in the "red noise regime." The two methods produce SF slope distributions peaking at $\gamma = 0.55 \pm 0.08$ and 0.52 ± 0.06, respectively, which are slightly steeper than but consistent with a single DRW model (Figure 12). The covariance function (or the autocorrelation function) of the underlying stochastic process leading to variability is well described by the exponential $\exp(-(\Delta t/\tau)^{\beta})$, where $\beta = 2\gamma$ is in the vicinity of unity. Of the theoretically tested scenarios on the origin of variability with the available predictions on SFs, the most viable is the model of accretion disk instabilities with $\beta \approx 0.9$ (Kawaguchi et al. 1998). Because SFs are model-independent and hence "raw" measures of variability, in fact we provide, for the first time, the direct proof that AGN variability is akin to the DRW process. The caveat is, however, that other covariance functions not considered here may potentially lead to similar SF slopes. Kozłowski (2016) simulated both DRW and non-DRW AGN light curves, modeled them as the DRW process, and found that it is so flexible that returns equally good fits for both DRW and non-DRW stochastic processes. Hence, light curve modeling with DRW cannot be used as a proof for the variability being caused by the DRW process. In fact, in Kozłowski et al. (2010a) we had already shown that DRW can correctly model deterministic processes (i.e., nonstochastic) such as pulsating variable stars, where the timescale is identified with the variability period. Whether DRW is the only underlying process we observe is not clear. While the distribution of the SF slopes peaks at $\gamma \approx 0.5$, as expected for DRW, we also observe a weak steepening of the SF for bright AGNs (with the lack of correlation with the black hole mass or Eddington ratio). This effect is observed both in the SF modeling with freed slopes, but also when fixing $\gamma = 0.5$ as an increase of the decorrelation timescale with the increasing luminosity.

The typical SF amplitude at one year is $\mathrm{SF}_0 = 0.22 \pm 0.06$ mag, but an SPL fit overestimates this value because the true SF is already in the pink noise regime (turning over). The true value of the SF amplitude at one year is $SF_0 = 0.20$ mag, while the asymptotic variability is $\mathrm{SF}_\infty = 0.25 \pm 0.06$ mag. The asymptotic variability SF_∞ is correlated with the black hole mass (with an SPL index of 0.29 ± 0.04) and strongly anticorrelated with the luminosity (with an SPL index of -0.35 ± 0.05).

The distribution of decorrelation timescales differs from a Gaussian but can be approximated with $\tau = 0.97 \pm 0.46$ year (median $\tau = 0.98$ year, $\sigma_\tau = 0.41$ year obtained from IQR), despite a weak prior on τ to be 1.5 year added in the minimization procedure. The timescale is correlated with the black hole mass (with an SPL index of 0.38 ± 0.15) and does not depend on luminosity (an SPL index of -0.05 ± 0.17).

AGN variability can be further studied with longer, already existing or near-future data sets. There exists a data set of about 800 AGNs lying behind the Magellanic Clouds and discovered mostly by the Magellanic Quasars Survey (Kozłowski et al. 2013). They have been observed for nearly two decades by the OGLE survey (Udalski et al. 2015) and should provide further clues on the decorrelation timescale, as the rest-frame timescales will be $\Delta t > 6$ years. The advantage of these light curves is twofold: (1) they have a cadence of a few days and were obtained with a single telescope with nearly identical detector setup (filters, pixel scale) for different phases of the OGLE survey, which will ease and make robust the data analysis; (2) they are viewed through the Magellanic Clouds, so a large number of constant stars are available to correctly estimate the photometric noise. The *GAIA* satellite (Perryman et al. 2001), collecting data since mid-2014, is planned to last five years, and it will scan the sky a median of 70 times. It will provide a usable photometry for broad $g < 20$ mag for a billion objects, including hundreds of thousands of AGNs. While these data will weakly probe the SF at the decorrelation timescale (they simply would have to be longer), they will enable a study of the shape of the ACF from ensembles based on a really large number of AGNs. The Large Synoptic Survey Telescope (Ivezic et al. 2008) will be a 10-year-long, deep $r < 24.5$ AB mag survey scanning three-quarters of the sky in six optical/infrared filters from the southern hemisphere. Although the typical cadence per filter will be only a month, it will provide excellent data sets in six filters to study AGN SF/ACFs and their dependence on the physical parameters of AGNs.

We are grateful to Chris Kochanek, Andrzej Udalski, and the anonymous referee for reading the manuscript and providing us with comments that improved the flow and clarity of the presented arguments. SK acknowledges the financial support of the Polish National Science Center through the OPUS grant number 2014/15/B/ST9/00093 and MAESTRO grant number 2014/14/A/ST9/00121.

REFERENCES

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Baganoff, F. K., & Malkan, M. A. 1995, ApJL, 444, L13
Bauer, A., Baltay, C., Coppi, P., et al. 2009, ApJ, 696, 1241
Blackburne, J. A., Pooley, D., Rappaport, S., & Schechter, P. L. 2011, ApJ, 729, 34
Butler, N. R., & Bloom, J. S. 2011, AJ, 141, 93
Chen, X., & Taam, R. E. 1995, ApJ, 441, 354
Choi, Y., Gibson, R. R., Becker, A. C., et al. 2014, ApJ, 782, 37
Czerny, B., Jaroszynski, M., & Czerny, M. 1994, MNRAS, 268, 135
Dai, X., Kochanek, C. S., Chartas, G., et al. 2010, ApJ, 709, 278
di Clemente, A., Giallongo, E., Natali, G., Trevese, D., & Vagnetti, F. 1996, ApJ, 463, 466
de Vries, W. H., Becker, R. H., White, R. L., & Loomis, C. 2005, AJ, 129, 615
Emmanoulopoulos, D., McHardy, I. M., & Uttley, P. 2010, MNRAS, 404, 931
Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748
Gould, A. 2003, arXiv:astro-ph/0310577
Graham, M. J., Djorgovski, S. G., Drake, A. J., et al. 2014, MNRAS, 439, 703
Greenstein, J. L. 1963, Natur, 197, 1041
Hawkins, M. R. S. 1993, Natur, 366, 242
Hook, I. M., McMahon, R. G., Boyle, B. J., & Irwin, M. J. 1994, MNRAS, 268, 305
Ivezić, Ž., Smith, J. A., Miknaitis, G., et al. 2007, AJ, 134, 973
Ivezic, Z., Tyson, J. A., Abel, B., et al. 2008, arXiv:0805.2366
Kasliwal, V. P., Vogeley, M. S., & Richards, G. T. 2015a, MNRAS, 451, 4328
Kasliwal, V. P., Vogeley, M. S., Richards, G. T., Williams, J., & Carini, M. T. 2015b, MNRAS, 453, 2075
Kawaguchi, T., Mineshige, S., Umemura, M., & Turner, E. L. 1998, ApJ, 504, 671
Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009, ApJ, 698, 895
Kelly, B. C., Becker, A. C., Sobolewska, M., Siemiginowska, A., & Uttley, P. 2014, ApJ, 788, 33
Kelly, B. C., Sobolewska, M., & Siemiginowska, A. 2011, ApJ, 730, 52

Kochanek, C. S. 2004, ApJ, 605, 58
Kozłowski, S., Kochanek, C. S., Stern, D., et al. 2010b, ApJ, 716, 530
Kozłowski, S., Kochanek, C. S., Udalski, A., et al. 2010a, ApJ, 708, 927
Kozłowski, S., Kochanek, C. S., Ashby, M. L. N., et al. 2016, ApJ, 817, 119
Kozłowski, S., Onken, C. A., Kochanek, C. S., et al. 2013, ApJ, 775, 92
Kozłowski, S. 2016, MNRAS, 459, 2787
MacLeod, C. L., Brooks, K., Ivezić, Ž., et al. 2011, ApJ, 728, 26
MacLeod, C. L., Ivezić, Ž., Kochanek, C. S., et al. 2010, ApJ, 721, 1014
MacLeod, C. L., Ivezić, Ž., Sesar, B., et al. 2012, ApJ, 753, 106
MacLeod, C. L., Ivezić, Ž., Sesar, B., et al. 2014, ApJ, 782, 119
Matthews, T. A., & Sandage, A. R. 1963, ApJ, 138, 30
Morgan, C. W., Kochanek, C. S., Morgan, N. D., & Falco, E. E. 2010, ApJ, 712, 1129
Mushotzky, R. F., Edelson, R., Baumgartner, W., & Gandhi, P. 2011, ApJL, 743, L12
Ozernoi, L. M., & Chertoprud, V. E. 1966, SvRP, 10, 15
Pancoast, A., Brewer, B. J., & Treu, T. 2014, MNRAS, 445, 3055
Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339
Press, W. H., Rybicki, G. B., & Hewitt, J. N. 1992a, ApJ, 385, 404
Press, W. H., Rybicki, G. B., & Hewitt, J. N. 1992b, ApJ, 385, 416
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992c, Numerical Recipes in FORTRAN. The Art of Scientific Computing (2nd ed.; Cambridge: Cambridge Univ. Press)
Rees, M. J. 1984, ARA&A, 22, 471
Rokaki, E., Collin-Souffrin, S., & Magnan, C. 1993, A&A, 272, 8
Ruan, J. J., Anderson, S. F., Dexter, J., & Agol, E. 2014, ApJ, 783, 105
Ruan, J. J., Anderson, S. F., MacLeod, C. L., et al. 2012, ApJ, 760, 51
Rybicki, G. B., & Press, W. H. 1992, ApJ, 398, 169
Scargle, J. D. 1981, ApJS, 45, 1
Scargle, J. D. 1982, ApJ, 263, 835
Scargle, J. D. 1989, ApJ, 343, 874
Schmidt, M. 1963, Natur, 197, 1040
Schmidt, K. B., Marshall, P. J., Rix, H.-W., et al. 2010a, ApJ, 714, 1194
Schmidt, K. B., Marshall, P. J., Rix, H.-W., et al. 2010b, ApJ, 721, 1941
Schneider, D. P., Hall, P. B., Richards, G. T., et al. 2007, AJ, 134, 102
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Shakura, N. I., & Sunyaev, R. A. 1976, MNRAS, 175, 613
Shen, Y., Greene, J. E., Strauss, M. A., Richards, G. T., & Schneider, D. P. 2008, ApJ, 680, 169
Simm, T., Salvato, M., Saglia, R., et al. 2016, A&A, 585, A129
Simonetti, J. H., Cordes, J. M., & Spangler, S. R. 1984, ApJ, 284, 126
Skowron, J., Udalski, A., Kozłowski, S., et al. 2016, AcA, 66, 1
Sumi, T., Woźniak, P. R., Eyer, L., et al. 2005, MNRAS, 356, 331
Smith, H. J., & Hoffleit, D. 1963, Natur, 198, 650
Udalski, A., Szymański, M. K., & Szymański, G. 2015, AcA, 65, 1
Uhlenbeck, G. E., & Ornstein, L. S. 1930, PhRv, 36, 823
Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, AJ, 122, 549
Vanden Berk, D. E., Wilhite, B. C., Kron, R. G., et al. 2004, ApJ, 601, 692
Wilhite, B. C., Brunner, R. J., Grier, C. J., Schneider, D. P., & vanden Berk, D. E. 2008, MNRAS, 383, 1232
Zu, Y., Kochanek, C. S., & Peterson, B. M. 2011, ApJ, 735, 80
Zu, Y., Kochanek, C. S., Kozłowski, S., & Udalski, A. 2013a, ApJ, 765, 106
Zu, Y., Kochanek, C. S., Kozłowski, S., & Peterson, B. M. 2016, ApJ, 819, 122

The Astrophysical Journal, 847:144 (3pp), 2017 October 1 https://doi.org/10.3847/1538-4357/aa89ee

Erratum: "Revisiting Stochastic Variability of AGNs with Structure Functions" (2016, ApJ, 826, 118)

Szymon Kozłowski
Warsaw University Observatory Al. Ujazdowskie 400–478 Warszawa, Poland; simkoz@astrouw.edu.pl
Received 2017 August 4; revised 2017 August 30; published 2017 October 3

1. Erratum

This erratum is meant to fix or clarify three issues in the paper by Kozłowski (2016).

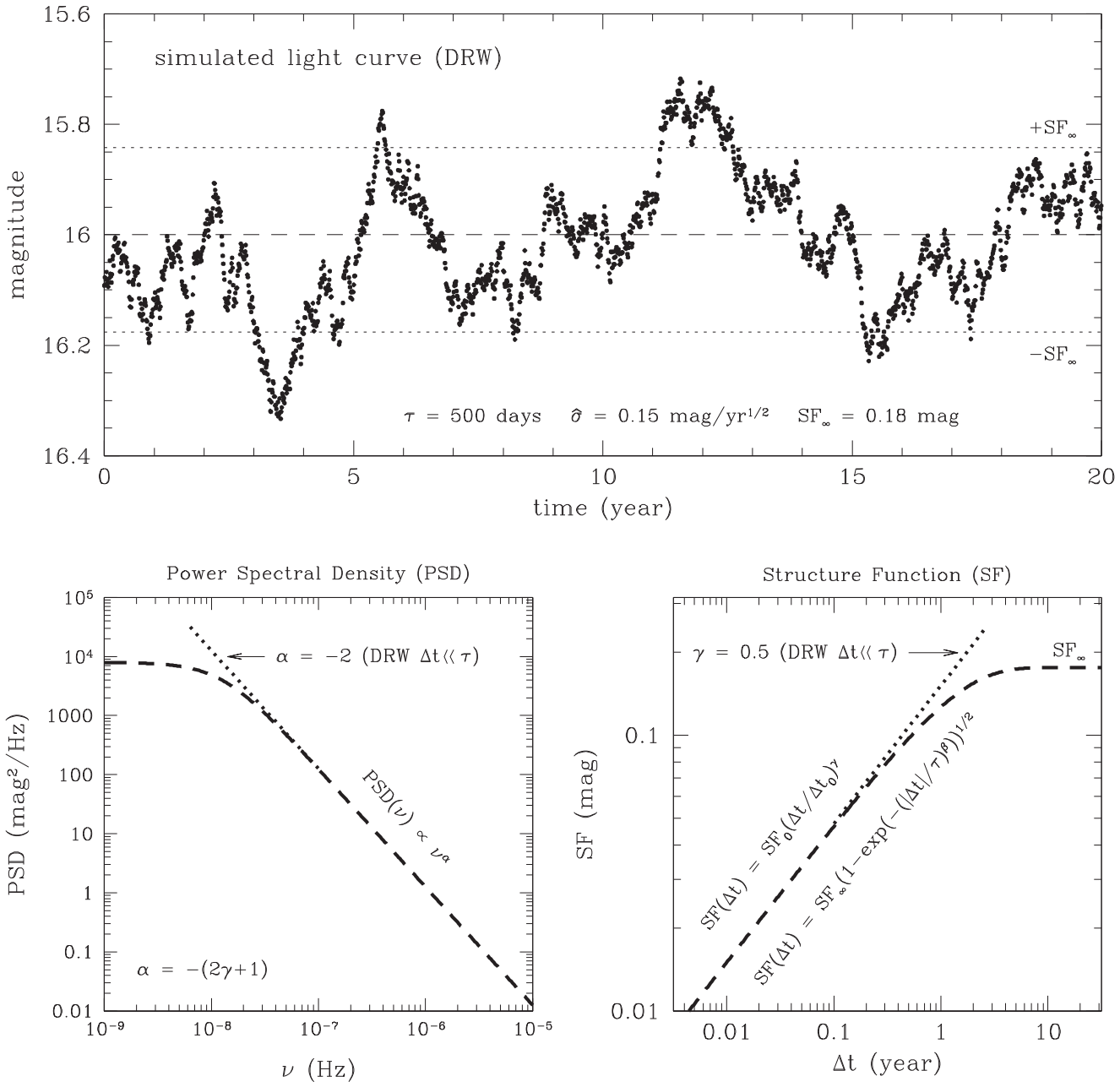

Figure 1. Presentation of basic active galactic nuclei (AGN) variability concepts and measures. Top panel: an idealized simulated AGN light curve using damped random walk (DRW) is shown. It can be either directly fit by the DRW (or continuous-time autoregressive moving average, CARMA) model and return the model parameters or studied via the PSD (bottom left panel) or SF (bottom right panel) analysis. Each panel also presents basic variability features related to these AGN variability measures.

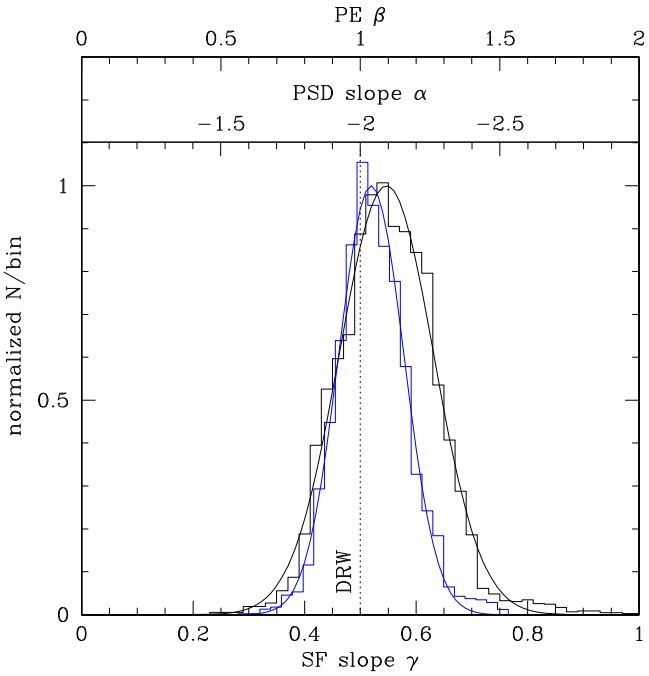

Figure 12. Histograms, weighted with the number of AGNs in a bin and normalized so the Gaussian fits peak at 1, of SF slopes γ (PE β and PSD slope α on top) from fitting the full SF function (black) and a single power law (SPL) SF for $4 < \Delta t < 365$ days (blue) to the redshift–absolute magnitude bins (corrected for biases). The histograms can be modeled relatively well as Gaussians with the mean values $\gamma = 0.55$ and 0.52, and dispersions $\sigma_\gamma = 0.08$ and 0.06, respectively. The DRW model ($\gamma = 0.5$) is marked with the vertical dotted line.

(1) In Section 2.5, I incorrectly assumed, instead of thoroughly checking, that for the power exponential (PE) auto-correlation function (ACF) of the form

$$\mathrm{ACF}(\Delta t) = e^{-(\frac{|\Delta t|}{\tau})^{\beta}}, \tag{27}$$

when $\beta \to 0$, the PE ACF turns into the white noise ACF, which is obviously not true, as then $\mathrm{ACF}(\Delta t) = 0.37$ for all Δt. The white noise ACF is 1 for $\Delta t = 0$ or 0 for $\Delta t \neq 0$. In fact, for $\beta = 0$, the ACF from Equation (27) is not even a valid ACF function as it does not obey the basic ACF property because ACF for $\Delta t = 0$ must be equal to 1. From this false assumption, I claimed that for $\beta \to 0$, the power spectral density (PSD) power-law slope is $\alpha \to 0$, which is again not true, as it is $\alpha \to -1$. This can be checked by doing the inverse Fourier transform of the ACF to obtain PSD, so

$$\mathrm{PSD}(\nu) = \int_{-\infty}^{\infty} \mathrm{ACF}(\Delta t)\, e^{-2\pi i \nu \Delta t} d\Delta t. \tag{25}$$

The conversions between the power-law slope powers of the PSD (α), structure functions (SF; γ), and the PE (β) are

$$\alpha = -(\beta + 1), \tag{E1}$$

and since the SF slope $\gamma = 0.5\beta$, we have

$$\alpha = -(2\gamma + 1), \tag{E2}$$

which is exactly the equation as reported by e.g., Bauer et al. (2009). I correct Figures 1 and 12 accordingly. Because the PSD analyses were not used in this paper, this has no impact on the results presented in the paper.

(2) In Section 2.4, Equation (20) is presented in a "simplified" notation, where the IQR(Δt) means the interquartile range of the magnitude differences Δm of data points separated by Δt, and not IQR of plain Δt. Similarly, the IQR(n) means, the interquartile range of the magnitude differences Δm of data points separated by $\Delta t < 2$ days, are hence regarded as the photometric noise n.

(3) In Section 4.2, both discussions and conclusions on the impact of the data length or cadence on the measured DRW parameters are too simplistic. I have thoroughly studied these issues in a subsequent paper and the reader is kindly asked to refer to that paper (Kozłowski 2017).

S.K. acknowledges the financial support of the Polish National Science Center through the OPUS grant number 2014/15/B/ST9/00093 and MAESTRO grant number 2014/14/A/ST9/00121.

ORCID iDs

Szymon Kozłowski https://orcid.org/0000-0003-4084-880X

References

Bauer, A., Baltay, C., Coppi, P., et al. 2009, ApJ, 696, 1241
Kozłowski, S. 2016, ApJ, 826, 118
Kozłowski, S. 2017, A&A, 597, A128