Light Curve Simulation

following Timmer & Koenig, Emmanoulopoulos et al. and Max-Moerbeck et al.

Lena Linhoff

Simulated light curves are used to estimate the confidence interval of the cross-correlation of two light curves obtained from the same source at different wavebands. To obtain the correct confidence intervals, the simulated light curves have to be uncorrelated but must feature the same Power Spectral Density (PSD) and Probability Distribution Function (PDF) as the original light curve. The approach used here follows the methods proposed by Emmanoulopoulos, McHardy, and Papadakis 2013 and Timmer and Koenig 1995 with some additional features proposed by Max-Moerbeck et al. 2014.

The method proposed by Timmer and Koenig results in artificial light curves following the original PSD only. Emmanoulopoulos, McHardy, and Papadakis extend this approach to an algorithm for generating light curves with the correct PSD *and* PDF. Furthermore, Max-Moerbeck et al. suggest some additional steps for estimating the underlying PSD from the periodogram of the original light curve.

The terms PDF, PSD and periodogram are defined as follows:

- PSD: Continuous Fourier spectrum of the true function (in our case: flux over time). In case of Active Galactic Nuclei (AGN) light curves, the PSD follows a $1/f^{\alpha}$ spectrum. The true distribution can never be measured since we are limited by the discrete sampling of the time series. For astronomy data, this is always the case because we usually have a very sparse time sampling with one data point per day, week, or even months, often interrupted by large gaps.
- Periodogram: The discrete Fourier spectrum of the time series sampled from the true function. The better the time resolution, the better the PSD can be estimated by the periodogram.
- PDF: Distribution of the absolute (flux) values of the time series.

1 Discrete Fourier Transform and Periodogram

Following Emmanoulopoulos, McHardy, and Papadakis, the periodogram is calculated based on the Discrete Fourier Transform (DFT) of an evenly sampled[1] time series $x(t_k)$ of length N and bin width t_{bin}:

[1]Light curves are never evenly sampled, which I will address in a minute.

$$\mathrm{DFT}(j) = \sum_{k=1}^{N} x(t_k) \exp\left(\frac{2\pi \mathrm{i}(k-1)j}{N}\right) \tag{1}$$

$$\text{for } j = 0 ... N-1 \ . \tag{2}$$

The DFT results in a series of complex numbers. The corresponding Fourier frequencies for even and odd N are defined as:

***N* even:**

$$f_j = \begin{cases} f_0 = 0 & \text{for } j = 0 \\ f_j^+ = j/Nt_{\mathrm{bin}} & \text{for } j = 1 \ldots \frac{N}{2} - 1 \\ f_{N/2} = f_{\mathrm{Ny}} = 1/2t_{\mathrm{bin}} & \text{for } j = \frac{N}{2} \\ f_j^- = -(N-j)/Nt_{\mathrm{bin}} & \text{for } j = \frac{N}{2} + 1 \ldots N-1 \end{cases} \tag{3}$$

***N* odd:**

$$f_j = \begin{cases} f_0 = 0 & \text{for } j = 0 \\ f_j^+ = j/Nt_{\mathrm{bin}} & \text{for } j = 1 \ldots \frac{N-1}{2} \\ f_j^- = -(N-j)/Nt_{\mathrm{bin}} & \text{for } j = \frac{N+1}{2} \ldots N-1 \end{cases} \tag{4}$$

$$\text{The Nyquist frequency } f_{N/2} \text{ does not exist for odd } N. \tag{5}$$

f^+ and f^- denote the positive and the nagative Fourier components, respectively.

Amplitude $\mathcal{A}$ and phase Φ of the complex numbers are defined as:

$$\mathcal{A}_j = \frac{1}{N} \tag{6}$$

and

$$\Phi_j = \arg[\mathrm{DFT}(j)] = \arctan\{\mathrm{Im}[\mathrm{DFT}(j)], \mathrm{Re}[\mathrm{DFT}(j)]\} \ . \tag{7}$$

The periodogram $P(f_j)$ can now be calculated as the squared amplitude of the j-th component:

$$P(f_j) = \mathcal{A}_j^2 = \frac{1}{N^2}\{\mathrm{Re}[\mathrm{DFT}(j)]^2 + \mathrm{Im}[\mathrm{DFT}(j)]^2\} \tag{8}$$

$$\text{for } j = 0 \ldots N-1 \ . \tag{9}$$

In our case, the Fourier transformed time series (the light curve) contains only real valued numbers, $x(t_k) \in \mathbb{R}$. Therefore, the positive DFT components are the complex conjugated negative DFT components and their amplitudes are equal:

$$\mathrm{DFT}(j^-) = [\mathrm{DFT}(j^+)]^* \tag{10}$$

$$\mathcal{A}_{j+} = \mathcal{A}_{j-} \ . \tag{11}$$

The periodogram of a real valued time series can be written as

$$P(f_j) = \mathcal{A}_j^2 = \frac{2}{N^2}\{\mathrm{Re}[\mathrm{DFT}(j)]^2 + \mathrm{Im}[\mathrm{DFT}(j)]^2\} \tag{12}$$

$$\text{with } f_j = \frac{j}{N t_{\mathrm{bin}}} \tag{13}$$

$$j = 0 \ldots \frac{N}{2} \qquad \text{for even } N \tag{14}$$

$$j = 0 \ldots \frac{N-1}{2} \qquad \text{for odd } N\ . \tag{15}$$

Following Emmanoulopoulos, McHardy, and Papadakis, the periodogram is normalized with the fractional root mean square factor $N t_{\mathrm{bin}}/\mu^2$ as proposed by Vaughan et al. 2003. μ denotes the mean of the original time series values. The periodogram is finally defined as:

$$P(f_j) = \frac{2 t_{\mathrm{bin}}}{\mu^2 N^2}\{\mathrm{Re}[\mathrm{DFT}(j)]^2 + \mathrm{Im}[\mathrm{DFT}(j)]^2\}\ . \tag{16}$$

1.1 Sampling Effects

As stated at the beginning, the time series that is Fourier transformed has to be evenly sampled, which is almost never the case for astronomical light curves. The light curves are therefore interpolated and re-sampled to a user defined bin width t_{bin}, which should be in the same range as the original sampling. Max-Moerbeck et al. show that the periodogram of a re-binned evenly sampled light curve represents the underlying PSD of the light curve better as the periodogram of the unevenly sampled data. Note that we can never use the interpolated light curves to derive any statement of the absolute flux values during the times between the measurements! The interpolation just reduces the impact of the uneven time sampling on the periodogram. This effect produces a bigger difference between the PSD and the periodogram than the missing values in the light curve, which is why interpolating a light curve is fine for the PSD estimation.

Furthermore, there are two effects we have to be aware of if we work with a sampled time series of finite length: *Red Noise Leakage* and *Aliasing*.

Red noise leakage is caused by the finite length of the light curve, when power that is originally deposited at wavelengths longer than the observation time is transferred to higher frequencies. This effect has an impact on the slope of the periodogram. Based on the original shape and slope of the underlying PSD, this impact can vary. For a PSD with a spectral slope of $\alpha < 1.5$ the red noise leakage is usually negligible Vaughan et al. 2003.

Aliasing refers to the transfer of power from higher frequencies to lower frequencies and is a result of the discrete sampling, which restricts the time resolution (see Kirchner 2005 for further details).

To reduce this effects for the PSD estimation, Max-Moerbeck et al. 2014 propose to convolute the original light curve with a Hanning window[2]. The Hanning window is

[2]Named after Julius von Hann, also called Hann function. Not to be confused with the Hamming function, another window function named after Richard Hamming.

defined as

$$w_{\text{Hanning}}(t) = \begin{cases} \cos^2\left(\pi \frac{t-T/2}{T}\right), & 0 \leq t \leq T \\ 0, & t > T \end{cases} \tag{17}$$

with the length T of the light curve.

Summing up the prequel for the PSD estimation we have to:

1. Apply a Hanning window $w_{\text{Hanning}}(t)$ to the original data $x(t_k)$.
2. Interpolate and re-bin the data for an evenly sampled light curve.
3. Calculate the $\text{DFT}(j)$ from the re-binned light curve using Equation 1.
4. Calculate the periodogram $P(f_j)$ from the Fourier components using Equation 16.

For the PDF estimation, the interpolated flux (without applying the window function!) must be scaled to values between 0 and 10. From these values the histogram for the PDF estimation can be calculated.

2 PSD and PDF Estimation

Before artificial light curves can be simulated that follow a certain PSD and PDF, these functions must be estimated from the original light curve. In the last section, the periodogram $P(f_j)$ was obtained from the original data. We can now fit a PSD model $\mathcal{P}(f_j)$ to this periodogram. Furthermore, the PDF of the original light curve has to be fitted. Both fits are executed using a maximum likelihood estimation.

2.1 PSD Likelihood

The PSD model used in Emmanoulopoulos, McHardy, and Papadakis 2013 is a bending power-law modyfied with an additional constant Poisson noise c:

$$\mathcal{P}(f;\boldsymbol{\gamma},c) = \frac{Af^{-\alpha_{\text{low}}}}{1+(f/f_{\text{bend}})^{\alpha_{\text{high}}-\alpha_{\text{low}}}} + c \tag{18}$$

with $\boldsymbol{\gamma} = (A, f_{\text{bend}}, \alpha_{\text{low}}, \alpha_{\text{high}})$ representing the model parameters. The Poisson noise c is added later to the model and is not a fit parameter.

To obtain the likelihood function, we assume the components of the periodogram $P(f_j)$ to be asymptotically χ^2 distributed around the true $\mathcal{P}(f_j)$:

$$P(f_j) = \begin{cases} \frac{1}{2}\chi_2^2\mathcal{P}(f_j), & j = 1 \ldots N/2-1 \text{ for even } N \\ \frac{1}{2}\chi_1^2\mathcal{P}(f_{\text{Ny}}), & j = N/2 \text{ for even } N \\ \frac{1}{2}\chi_2^2\mathcal{P}(f_j), & j = 1 \ldots (N-1)/2 \text{ for odd } N\,. \end{cases} \tag{19}$$

χ_ν^2 denotes the χ^2 distribution with ν degrees of freedom.

From these distributions we can express probability to obtain a single periodogram component $P(f_j)$ if the underlying PSD is given as $\mathcal{P}(f_j;\boldsymbol{\gamma})$:

$$\lambda\left[P(f_j)|\mathcal{P}(f_j;\boldsymbol{\gamma})\right] = \begin{cases} \mathit{\Gamma}\left[\frac{\nu=2}{2},\mathcal{P}(f_j;\boldsymbol{\gamma})\right] & j = 1 \ldots N/2-1 \text{ for even } N \\ \mathit{\Gamma}\left[\frac{\nu=1}{2},\mathcal{P}(f_j;\boldsymbol{\gamma})\right] & j = N/2 \text{ for even } N \\ \mathit{\Gamma}\left[\frac{\nu=2}{2},\mathcal{P}(f_j;\boldsymbol{\gamma})\right] & j = 1 \ldots (N-1)/2 \text{ for odd } N\ . \end{cases} \tag{20}$$

The unbinned likelihood for obtaining a full periodogram from a certain PSD is then given as

$$\mathcal{L} = \prod_{i=1}^{\substack{N/2 \text{ (even } N) \\ (N-1)/2 \text{ (odd } N)}} \lambda_j\left[P(f_j)|\mathcal{P}(f_j;\boldsymbol{\gamma})\right]\ . \tag{21}$$

The best model estimation for $\boldsymbol{\gamma}$ is now obtained by minimizing the negative logarithmic likelihood function.

2.2 PDF Likelihood

For the PDF estimation, the flux values are re-scaled to values between 0 and 10 which is a suitable range for the standard gamma and log-norm distributions. The normed flux is then binned in a histogram to obtain the flux distribution, which has to be fitted.

The PDF of an AGN light curve can often be modeled as a superposition of two distributions representing two states of activity: The low state and the flaring activity. For this work, the PDF is modeled with a gamma and a log-normal distribution:

$$\mathrm{PDF}(x;\boldsymbol{\eta}) = w_1\gamma(x,a) + w_2\mathcal{N}_{\log}(x,s,\mu,\sigma) \tag{22}$$

$$= w_1 \cdot \frac{x^{a-1}e^{-x}}{\mathit{\Gamma}(a)} + w_2 \cdot \frac{1}{s(x-\mu)\sqrt{2\pi}} \exp\left(-\frac{\log^2(\frac{x-\mu}{\sigma})}{2s}\right) \tag{23}$$

with the weights $w_{1,2}$ so that $w_1 + w_2 = 1$ and the fit parameters $\boldsymbol{\eta} = (a, s, \mu, \sigma)$. The unbinned likelihood is given as

$$\mathcal{L} = \prod_{k=1}^{N} \mathrm{PDF}(x_k|\boldsymbol{\eta})\ . \tag{24}$$

For a *binned* likelihood approach, the flux values of the light curve are binned in n_{bin} bins of the size $\mathit{\Delta}x$. The bin value can now be estimated as the integral of the PDF in the corresponding bin edges. The integral of the PDF is given by the Cumulative Distribution Function (CDF), in this case the weighted sum of the CDF of the gamma and the log-norm distributions as used in Equation 22. To estimate the value in a bin, the CDF must be evaluated at the bin edges. The binned likelihood is then given as

$$\mathcal{L} = \prod_{i=1}^{n_{\mathrm{bins}}} \mathrm{CDF}((i+1)\mathit{\Delta}x|\boldsymbol{\eta}) - \mathrm{CDF}(i\mathit{\Delta}x|\boldsymbol{\eta})\ . \tag{25}$$

If the binned or the unbinned approach is used depends on the amount of data. For an unbinned fit, the computation time scales with the number of flux points, which is why it makes sense to use a binned fit for long light curves. For light curves with few flux points, a binned likelihood fit might fail and an unbinned approach should be chosen.

Having the likelihood defined, we have to take care of the maximization, respectively, the minimization of the logarithmic likelihood. For this work, the minimization is performed with the `minuit` **James:1975dr** minimizer, which is implemented as `Python` package `iminuit` **iminuit**.

3 Timmer & König Algorithm

As already mentioned, Timmer and Koenig 1995 developed a method to generate artificial light curves featuring the same PSD as the original light curve. This approach is also used by Emmanoulopoulos et al., which is why I will describe it briefly. To obtain a time series from a given PSD $\mathcal{P}(f_j)$, given Fourier frequencies f_j and length N_{TK} of the light curve, the following steps have to be performed:

1. Draw two normal distributed random numbers from $\mathcal{N}(\mu = 0, \sigma = 1)$ for $j = 0 \ldots N_{\mathrm{TK}}/2$ (even) or $j = 0 \ldots (N_{\mathrm{TK}} - 1)/2$ (odd) and multiply them by $\sqrt{\frac{1}{2}\mathcal{P}(f_j)}$. The resulting numbers are used as real and imaginary part of the complex numbers that form the DFT of the simulated light curve. If N_{TK} is even, $\mathrm{DFT}(f_{\mathrm{Ny}} = f_{N_{\mathrm{TK}}/2})$ is real valued and only one random number has to be sampled.

2. To create a real valued light curve, calculate the complex conjugates of the sampled complex numbers for $j = (N_{\mathrm{TK}}/2) + 1 \ldots N_{\mathrm{TK}}$ (even) or $j = (N_{\mathrm{TK}} + 1)/2 \ldots N_{\mathrm{TK}}$ (odd).

3. The simulated light curve can now be calculated as the inverse Fourier transform of the sampled components.

Analog to the PSD estimation, we have to take care of red noise leakage and aliasing when we sample a light curve from a given PSD. The PSD of the simulated light curve can be different from the given one because of these effects.

To avoid red noise leakage in the simulated data, we create a light curve that is much longer than the original light curve ($N_{\mathrm{TK}} = 100...1000 \cdot N$) and take a random sequence of length N from the simulated light curve. The aliasing can be taken into account by simulating a light curve with smaller time bins t_{bin} and re-bin the result afterwards. How strong the effect of red noise leakage and aliasing is in the individual case, depends on the features of the original light curve and its sampling. Anyhow, a user should always keep this effects in mind. Both approaches to correct for red noise leakage and aliasing are implemented in my code.

4 Emmanoulopoulos Loop

After the PSD and PDF of a light curve are successfully estimated, the actual light curve sampling can finally start. For this purpose, I follow the algorithm proposed by Emmanoulopoulos, McHardy, and Papadakis 2013.

1. Create an artificial light curve x_{TK} from a given PSD $\mathcal{P}(f_j)$ with the method by Timmer & König. The input PSD should not contain the Poisson noise (denoted

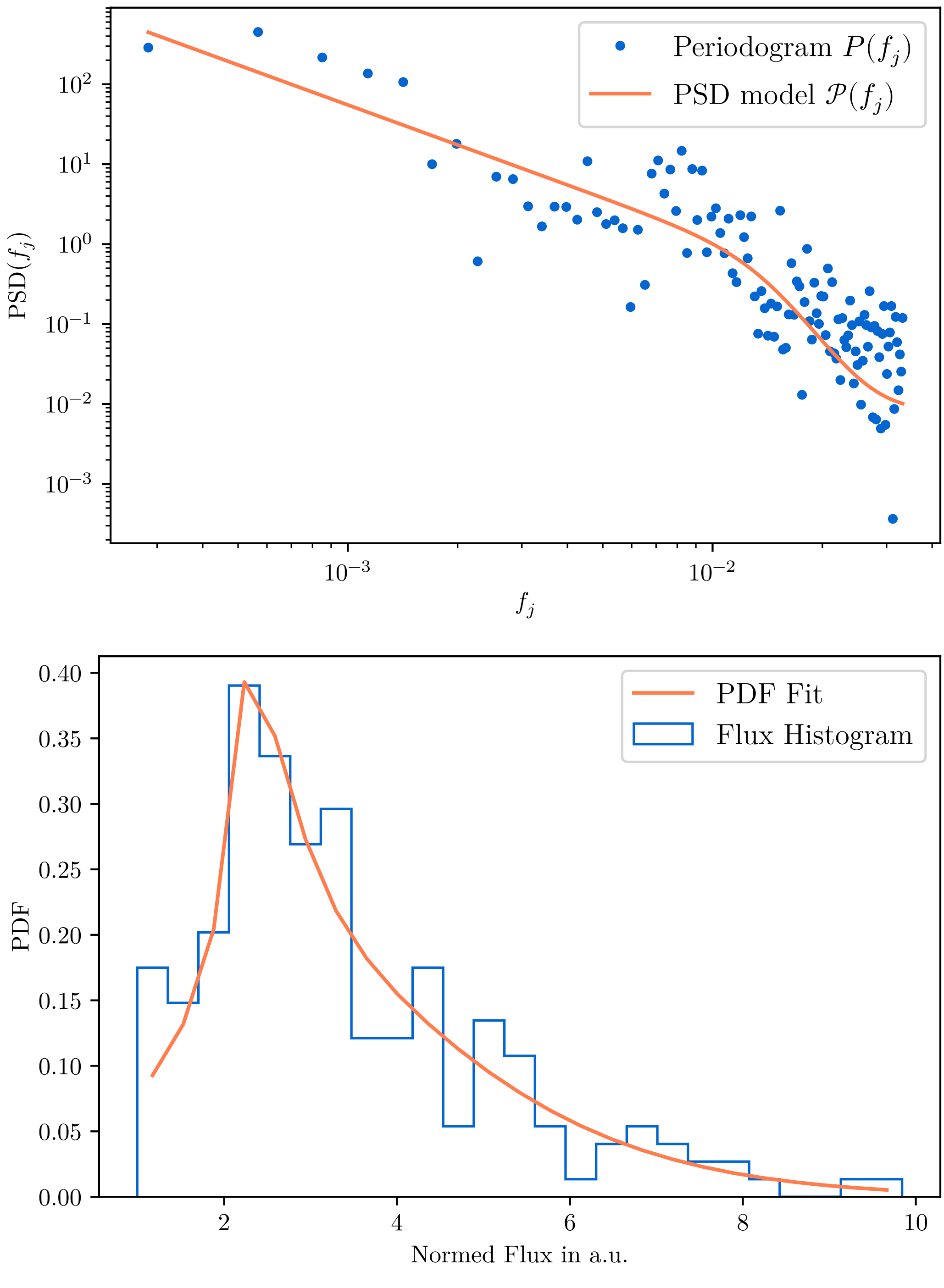

Figure 1: **Above:** Periodgram of a light curve (blue dots) and the likelihood fitted PSD for the model described in Equation 18. **Below:** Flux histogram with the likelihood estimation for the PDF. The model used here is described in Equation 22. The flux is scaled to values between 0 and 10.

as c in Equation 18), which is added later. This light curve features the correct PSD, but not the PDF. Calculate the DFT of x_{TK} and the amplitudes $\mathcal{A}_{\mathrm{TK}}$ and phases Φ_{TK} for the Fourier components.

2. Sample N values from the PDF, denoted as $x_{\mathrm{sim,\ 1}}$ and also calculate the DFT of $x_{\mathrm{sim,\ 1}}$, and $\mathcal{A}_{\mathrm{sim,1}}$ and $\Phi_{\mathrm{sim,1}}$ from the Fourier components.

3. **Spectral adjustment**: Create new Fourier components by combining $\mathcal{A}_{\mathrm{TK}}$ and $\Phi_{\mathrm{sim,1}}$ and calculate a new light curve x_{adjust} from the inverse Fourier transform of the comnined components. This light curve still features the correct PSD, but a different PDF.

4. **Amplitude adjustment**: Order the $x_{\mathrm{sim,\ 1}}$ based on the ranking of x_{adjust}: The highest value of x_{adjust} is replaced with the highest value of $x_{\mathrm{sim,\ 1}}$ and so on. The resulting light curve now features the desired PDF, but the PSD has changed.

5. Repeat the process from step 2, sampling $x_{\mathrm{sim,\ 2}}$ and replacing the values from $x_{\mathrm{sim,\ 1}}$ with $x_{\mathrm{sim,\ 2}}$. This is done several times until $x_{\mathrm{sim,\ k}}$ does not change anymore and the $P_k(f_j)$ follows the given $\mathcal{P}(f_j)$.

After a time series x is sampled this way, the Poisson noise has to be added to the light curve to take the count statistics of the detector into account. The light curve becomes

$$x_{\mathrm{sim,\ Poisson}}(t_i) \propto \frac{\mathrm{Poisson}[\mu = x_{\mathrm{sim}}(t_i)\Delta t]}{\Delta t} \qquad \text{for } i = 1 \dots N \tag{26}$$

with the time bin width Δt and the Poisson distribution with mean $x_{\mathrm{sim}}(t_i)$.

After this step is done, the final simulated light curve is obtained by interpolating and re-sampling the evenly sampled light curve $x_{\mathrm{sim,\ Poisson}}(t_i)$ in the same way as the original time series and re-scaling the flux values. For the purpose of cross-correlation studies, (several) thousand light curves have to be simulated to calculate the confidence levels.

References

[EMP13] D Emmanoulopoulos, IM McHardy, and IE Papadakis. “Generating Artificial Light curves: Revisited and Updated”. In: *Monthly Notices of the Royal Astronomical Society* 433.2 (2013), pp. 907–927. DOI: `10.1093/mnras/stt764`.

[Kir05] James W Kirchner. “Aliasing in 1/ f α Noise Spectra: Origins, Consequences, and Remedies”. In: *Physical Review E* 71.6 (2005), p. 066110. DOI: `10.1103/PhysRevE.71.066110`. URL: `http://seismo.berkeley.edu/~kirchner/reprints/2005_74_spectral_aliasing.pdf`.

[Max+14] W Max-Moerbeck et al. “A Method for the Estimation of the Significance of Cross-Correlations in Unevenly Sampled Red-Noise Time Series”. In: *Monthly Notices of the Royal Astronomical Society* 445.1 (2014), pp. 437–459. DOI: `10.1093/mnras/stu1707`.

[TK95] J Timmer and M Koenig. “On Generating Power Law Noise.” In: *Astronomy and Astrophysics* 300 (1995), p. 707. URL: `http://adsabs.harvard.edu/pdf/1995A%26A...300..707T`.

[Vau+03] S Vaughan et al. "On Characterizing the Variability Properties of X-Ray Light Curves from Active Galaxies". In: *Monthly Notices of the Royal Astronomical Society* 345.4 (2003), pp. 1271–1284. DOI: 10.1046/j.1365-2966.2003.07042.x.