

On the use of structure functions to study blazar variability: caveats and problems

D. Emmanoulopoulos^{1*}, I. M. McHardy¹ and P. Uttley¹

¹ School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ

Accepted 2009 Month #. Received 2009 Month #; in original form 2009 Month #

ABSTRACT

The extensive use of the structure function (SF) in the field of blazar variability across the electromagnetic spectrum suggests that characteristic time-scales are embedded in the light curves of these objects. We argue that for blazar variability studies, the SF results are sometimes erroneously interpreted leading to misconceptions about the actual source properties. Based on extensive simulations we caution that spurious breaks will appear in the SFs of almost all light curves, even though these light curves may contain no intrinsic characteristic time-scales. i.e. having a featureless underlying power-spectral-density (PSD). We show that the time-scales of the spurious SF-breaks depend mainly on the length of the artificial data set and also on the character of the variability i.e. the shape of the PSD.

The SF is often invoked in the framework of shot-noise models to determine the temporal properties of individual shots. We caution that although the SF may be fitted to infer the shot parameters, the resultant shot-noise model is usually inconsistent with the observed PSD. As any model should fit the data in both the time and the frequency domain the shot-noise model, in these particular cases, can not be valid.

Moreover, we show that the lack of statistical independence between adjacent SF points, in the standard SF formulation, means that it is not possible to perform robust statistical model fitting following the commonly used least-squares fitting methodology. The latter yields uncertainties in the fitting parameters (i.e. slopes, breaks) that are far too small with respect to their true statistical scatter. Finally, it is also commonly thought that SFs are immune to the sampling problems, such as data gaps, which affects the estimators of the PSDs. However we show that SFs are also troubled by gaps which can induce artefacts.

Key words: methods: statistical – methods: numerical – methods: data analysis – galaxies: individual: Mrk 501 – galaxies: active – BL Lacertae objects: general

1 INTRODUCTION

The flux variation of Active Galactic Nuclei (AGN) is a phenomenon that can provide us with significant information about the physical properties of these sources. The most variable AGN are the blazars which exhibit dramatic flux variations across the electromagnetic spectrum. The fastest variations are observed in the X-ray and γ -ray bands on time-scales of hours or even minutes (Catanese et al. 1997; Maraschi et al. 1999; Aharonian et al. 2003, 2005a,b) whereas in the optical (Tosti et al. 1998; Stalin et al. 2005) and radio wavebands (Romero et al. 1997; Aller et al. 1999; Teräsranta et al. 2005) the temporal variations are seen on time-scales of days to weeks, up to years.

In the last fifty years several time-series analysis meth-

ods have been developed to study the variability properties of astronomical sources (see for a review Subba Rao et al. 1997). Linear and nonlinear analysis methods in the time and frequency domains, adjusted to the needs of astronomical data sets i.e. taking into account measurement errors and data-gaps, can describe adequately the time behaviour of AGN. Nevertheless, several authors, as we are going to discuss further, are convinced that more conventional, intuitive-tools, based on “running variance” computations are able to give robust results with respect to the underlying variability properties of the observed source. Although these methods have value in some cases, they can often give misleading results as we shall show in this paper.

One of the most extensively used tools in the field of blazar variability is the *structure function* (SF) (see e.g. Hughes et al. 1992) which measures the mean value of the flux-variance for measurements, $x(t)$, that are separated by

* E-mail: D.Emmanoulopoulos@soton.ac.uk

a given time interval, τ , where $SF(\tau) = \langle [x(t) - x(t + \tau)]^2 \rangle$. The SF is commonly characterized in terms of its slope β , where $SF(\tau) \propto \tau^\beta$.

Historically the SF has been used in the study of turbulent plasmas (Kolmogorov 1941a,b; Yu et al. 2003). It was introduced to astrophysics by radio astronomers studying slow scintillations in the interstellar medium (Rickett et al. 1984) following the methodologies of Prokhorov et al. (1975) and Coles & Frehlich (1982) on the subjects of laser propagation in turbulent media and atmospheric scintillation respectively. The first systematic description of the SF methodology, adjusted to the needs of astronomical data sets, was made by Simonetti et al. (1985) using as a reference the work of Rutman (1978) from the field of electrical and electronic engineering. Simonetti et al. used the SF to demonstrate that the time-series of flat- and steep-spectrum radio sources differ qualitatively. During the same period, Cordes & Downs (1985) estimate the SFs of 21 pulsars and Hjellming & Narayan (1986) derived the first quantitative SF results for the compact galactic radio source 1741-038. Then, Fiedler et al. (1987); Heeschen et al. (1987) also used the SF to study the variability properties of large extragalactic radio samples. Subsequently, the SF has been employed for the study of the timing properties of objects in higher energy bands. Bregman et al. (1988); Neugebauer et al. (1989); Quirrenbach et al. (1992); Hufnagel & Bregman (1992); Smith et al. (1993); Lainela & Valtaoja (1993); Heidt & Wagner (1996) whilst Lewis & Irwin (1996) and Blandford & Kundic (1997) applied the SF-analysis method to derive masses for microlenses. In the X-ray band, Brinkmann et al. (2000) applied the SF methodology to the ROSAT observations of PKS 2155-304 and they attributed the derived shot-noise variability to either the accelerating inner jet model or to accelerating particles at shocks travelling down the relativistic jet. Thereafter many X-ray variability many X-ray observations have employed SF analysis (e.g. Gliozzi et al. 2001; Brinkmann et al. 2001; Iyomoto & Makishima 2001; Zhang et al. 2002) in order to derive some short of characteristic time-scale.

SFs have been particularly commonly applied to observations of blazars. The *ASCA* “long-look” observations of Mrk 501, Mrk 421, and PKS 2155-304 (Takahashi et al. 2000; Tanihata et al. 2001, 2003) are, even now, some of the most uniformly sampled light curves in the X-ray regime. The main outcome of the above SF analysis is an apparent characteristic time-scale of ~ 1 day which appears to be in accordance with the duration of the shots existing in the internal shock particle acceleration scenario (Sikora et al. 2001; Spada et al. 2001).

In this work we show through extensive simulations that much of the existing literature on blazar SFs tends to misinterpret observed SF characteristics such as breaks and slopes as being real or physically meaningful, when often these are either artefacts intrinsic to SFs, or subject to much greater statistical variation than inferred from the commonly-used fitting procedures. In Section 2 we present the method that we are going to use in order to create artificial light curves lacking any sort of characteristic time-scales and in Section 3 we study the behaviour of the SF derived from the thoroughly studied *ASCA* data set of Mrk 501 (Tanihata et al. 2001, hereafter TAN01). For the same source we employ the long-look light curve of *All-Sky Monitor (ASM)*, onboard

Rossi X-ray Timing Explorer (RXTE), in order to study the effects of the data length on the position of the SF-break. Then, in Section 4 we study the timing properties in the Fourier domain of the shot-noise model that TAN01 have used in order to study the SF properties. In Section 5 we test the statistical robustness of the most commonly used fitting procedures which are employed in order to derive astrophysically interesting quantities from the SF. Finally, in Section 6 we study the sensitivity of SF to the presence of data-gaps.

2 LIGHT CURVE SIMULATIONS AND POWER SPECTRAL DENSITY

We simulate stationary light curves based on the procedure used by Timmer & Koenig (1995) which uses as an input the Power Spectral Density (PSD) function of the observed light curve and gives an ensemble of stochastic time-series produced from the same underlying PSD. This is the most robust method of mimicking the properties of a light curve since it takes into account correctly the intrinsic scatter in the power at a given frequency, described by a χ^2_2 , chi-squared distribution with two degrees of freedom (e.g. Priestley 1981), by randomizing both phases and amplitudes.

Throughout this paper we use the standard nomenclature, i.e. we refer to the periodogram as the modulus squared of the discrete Fourier transform of the particular data set under consideration whilst the PSD represents the mean of the true underlying distribution of variability power as a function of frequency (Priestley 1981; Vaughan et al. 2003). Since the periodogram is highly scattered around the PSD having a standard deviation of 100 per cent (Press et al. 1992), we use the binned logarithmic periodogram (Papadakis & Lawrence 1993; Vaughan 2005). If there are more than 20 samples in each geometric-mean frequency bin then their distribution will be Gaussian, thus producing a statistically sensible standard deviation.

Concerning the length of the simulated light curves some special caution should be taken with respect to the effect of *red-noise leak* i.e. the transfer of power from low to high frequencies by the lobes of the window function which can produce features such as slowly increasing or falling trends across the generated light curve (e.g. Priestley 1981). To take this effect into account, we extend the PSD to very low frequencies (i.e. long time-scales), in order to generate artificial light curves 50 times longer than the observed one. We then truncate the simulated data trains to the desired length by selecting a random segment having equal length to the real light curve under study.

In the case of blazar variability the underlying PSD can be very-well described by simple or broken-power-laws with indices $1 < \alpha < 2.5$ flattening at long time-scales depicting the physical fact that the variability amplitude can not increase forever (stationarity). For these kinds of physically stationary processes the periodogram can correctly describe the underlying PSD even for steeper slopes of -4 (or even -5).

Apart from red-noise leak, spectral representations are also affected by *aliasing effects*. Frequency components outside the frequency range covered by the data set, are

aliased into that range by the very act of discrete sampling (Press et al. 1992). Finally, irregular sampling produces a window function which smears the spectral estimates (Scargle 1989), causing deviations from the true underlying PSD. Over the years a number of methods have been derived to overcome these problems. In particular, we can estimate the underlying PSD by model fitting (e.g. Done et al. 1992; Uttley et al. 2002). This process is now commonly applied to observations of Seyfert galaxy X-ray variability (e.g. Markowitz et al. 2003; Uttley & McHardy 2005; McHardy et al. 2004, 2005, 2006, 2007) and produces reliable results. Similar simulation-based treatments of the SF results are missing from the blazar literature.

3 CAVEATS REGARDING THE STRUCTURE FUNCTION BREAK-TIME-SCALES

3.1 The *ASCA* data set of Mrk 501

To study the behaviour of the SF, we examine the *ASCA* data set of Mrk 501 (TAN01) in 2–10 keV, with a sample interval of 5678.3 sec. This is one of the three data sets (the other two are Mrk 421 and PKS 2155-304) in which TAN01, based on the SF analysis method (Fig. 1, left panel), find characteristic time-scales and suggest that these are a signature of the minimum time-scales of individual shots. This is one of the longest and most continuous data sets ever obtained for a blazar in the X-rays, covering a time span of 10 days.

Initially, we estimate the binned logarithmic periodogram of the data set which can be very well fitted by a simple featureless power-law with index $\alpha = -1.80 \pm 0.09$ ($\chi^2 = 3.80$ for 6 degrees of freedom with a null hypothesis probability of 0.70). There is no evidence for a break in the PSD.

We next produce 2000 artificial light curves having the same power-law PSD of index $\alpha = -1.80$ and the same length as the studied *ASCA* data set. For each artificial light curve we estimate the SF by employing exactly the same functional form of the SF as TAN01

$$SF(\tau) = \frac{1}{N(\tau)} \sum w(i)w(i+\tau)[f(i+\tau) - f(i)]^2 \quad (1)$$

where $N(\tau) = \sum w(i)w(i+\tau)$, τ is the separation time of the observations, and $w(i)$ is the weighting factor which is proportional to the data point $f(i)$ and inversely proportional to its error $\sigma_f(i)$. By an eye inspection, we note apparent breaks in almost all of the simulated SFs.

In our simulations we take into account the effect of the measurement errors in the SF-estimates by representing each measurement as a deviate drawn from a Gaussian distribution with mean and standard deviation equal to the measurement’s value and error respectively (200 times for each artificial light curve). We have to note that for this particular *ASCA* data set the measurement errors do not play a major role in the SF-estimates. If we ignore these errors the values of the SF-breaks and SF-slopes change only by ~ 1 per cent and ~ 0.5 per cent respectively.

Next, we determine the position of the SF-break (Fig. 1, right panel). To localize the break we produce an interpolated version of the SF and we find the abscissa τ_{br} of the first local maximum, which indicates the position

of the break. Even if a clear plateau is not formed after $\tau > \tau_{br}$, several authors consider the first “hump” as a signature of a genuine temporal property of the source (e.g. Takahashi et al. 2000; Gliozzi et al. 2001; Kataoka et al. 2001; Agudo et al. 2006; Fuhrmann et al. 2008).

Finally we note that after the first break point, the SF usually exhibits “wiggly-patterns” forming a plateau. These fluctuations indicate that the pairs $\{f(i+\tau), f(i)\}_{\tau > \tau_{br}}$ which are averaged within each time bin $\tau > \tau_{br}$, are linearly independent having zero linear correlation (Appendix A). In order to study the frequency of occurrence of these SF structures, we register the abscissas of all the local extrema occurring for $\tau > \tau_{br}$.

In the left panel of Fig. 2 we present with the filled grey area the distribution of the SF-breaks coming from our simulated light curves that have the same length and the same featureless PSD as the original *ASCA* light curve of Mrk 501. From an ensemble of 2000 simulated data sets, we get breaks from 1903 of them. To specify the position of the maximum in our histogram we fit the histogram entries with a Gaussian distribution. The results of the fit yield an amplitude of 279.55 samples, a mean value of 0.91 days and a standard deviation of 0.09 days¹. With the solid and dotted lines we present the distribution of the local extrema (maxima with solid line and minima with the dotted line) occurring after the first break. The occurrence times of the “wiggling” patterns are purely randomly distributed for $\tau > \tau_{br}$ following a uniform distribution.

These simulations show us clearly that stochastic data sets having the same length as the *ASCA* data set of Mrk 501 and the same featureless PSD, exhibit breaks in the SF around a day. Thus, the apparent break seen in the SF for Mrk 501 (Fig. 1, left panel) should not be associated with any sort of physically meaningful time-scale.

3.2 Longer time-series: The *ASM* data set of Mrk 501

Following Section 3.1, we extend our simulations to even longer time-scales to check how SF-break time-scales might be related to the length of a data set. Assuming that for long time-scales the variability properties of MRK 501 are well-represented by the same PSD describing the short-term variability (i.e. same slope and normalisation), we produce 2000 artificial light curves, each being 2000 days long. For every light curve we calculate the SF and we specify the position of the SF-breaks. In total 1893 light curves exhibit SF-breaks and the distribution of their position can be well represented by a Gaussian distribution having a mean 399.35 days and a standard deviation of 74.73 days (Fig. 2, right panel).

In order to compare our predicted break with the data we use the long-term *RXTE-ASM* light curve of Mrk 501² spanning from MJD 50100 to MJD 52100 and we estimate

¹ By fitting a log-normal distribution the results remain practically the same giving an amplitude of 254.33 samples, $\mu = -0.09$, and $\sigma = 0.10$, yielding a mean value of 0.92 days and a standard deviation of 0.09 days.

² We have obtained the light curve from the *ASM* light curve site of MIT: http://xte.mit.edu/ASM_lc.html

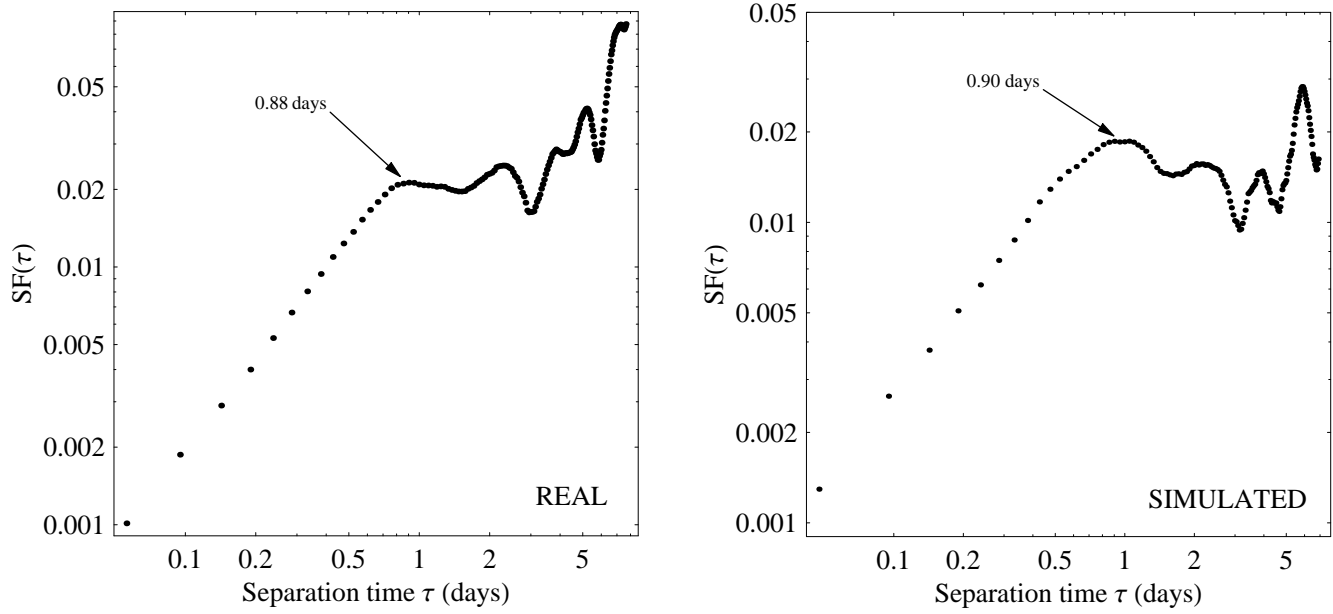


Figure 1. [Left panel] The SF (in logarithmic scale) for the *ASCA* data set of Mrk 501 as shown in panel (a) of figure 4 in TAN01. The arrow indicates the break time-scale which is around a day.

[Right panel] A randomly chosen SF (in logarithmic scale) from the ensemble of the 2000 artificial light curves which are produced from a featureless power-law PSD of index -1.80 with no characteristic time-scale, having the same length as the *ASCA* data set. There is a clear break around one day, similar to the one derived from the *ASCA* SF.

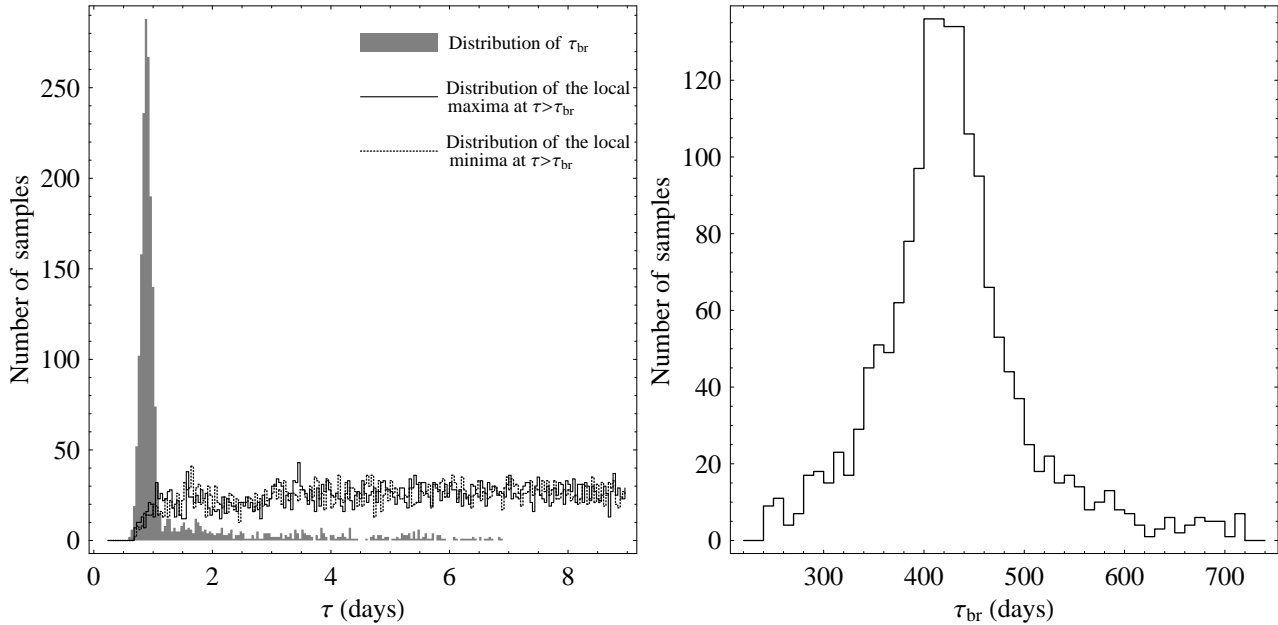


Figure 2. The distribution of the SF's breaks and wiggling features.

[Left panel] The grey area depicts the distribution of the SF-breaks coming from an ensemble of 2000 artificial light curves which are produced from a featureless power-law PSD of index -1.80 with no characteristic time-scale, for the case of the *ASCA* data of Mrk 501 (the histogram bins have a length of 0.04 days). Based on a Gaussian fit the mean value and the standard deviation of the distribution are 0.92 and 0.09 days respectively. The solid and dashed lines represent the distribution of the local maxima and local minima for $\tau > \tau_{br}$ mapping the positions of the wiggling features.

[Right panel] The distribution of the SF-breaks for the case in which the simulated light curves, produced by the same PSD as above, extend to 2000 days (the histogram bins have a length of 10 days). It has a mean value of 399.35 days and a standard deviation of 74.73 days.

its SF. The SF-break occurs around 402 days, something which is absolutely in accordance with the results from our simulated light curves of the same length which did not include of any sort of characteristic time-scale.

The aforementioned example reflects in a clear way that the SF deals only with the properties of the observed light curve ignoring the properties of the true underlying variability process. During astronomical observations we observe a source for a given time and from the observed data series we have to extract in a statistically robust way the properties of the underlying variability process. If the result that we obtain is a strong function of the length of a given data set then this can lead to a serious misunderstanding of the variability properties of the source.

Similar SF effects to those derived for the *ASCA* and *ASM* light curves of Mrk 501 should be expected for all similar light curves. By producing an artificial light curve 6000 time units (t.u.) long, from a featureless power-law PSD of slope -2, we estimate the normalized SF (NSF) (equation (A10)) (i.e. the SF divided by two times the value of the variance of the data set, see Appendix A) for the overall data train and for a 500 t.u. subset of it (Fig. 4). In this case we use the NSF, instead of the classical SF, for an easy and direct comparison of the position of the break, since the formed plateaus in both SFs are distributed around 2. The top right panel of Fig. 4 shows the behaviour of the NSF for these two data sets, coming from the same underlying variability process. The break in the NSF for the short data set occurs at around 40 days and for the overall light curve at 1600 days respectively. In this case, the fact that the break occurs at two different time points for two different durations of the data set which is fully described from the same and completely featureless PSD, indicates that the break does not reflect any physically interesting time property of the process itself.

For comparison purposes we also estimate the binned logarithmic periodograms for the two light curves discussed above. As we can see from the bottom left panel of (Fig. 4) both can very well reproduce the input PSD slope of -2.

3.3 SF-breaks and data set length

To quantify the relationship between the spurious SF-breaks, τ_{br} , and both the length and the PSD-slope of the data set, we perform a set of simulations. For each featureless power-law PSD (with index $\alpha = -1$, $\alpha = -1.5$, $\alpha = -2$, $\alpha = -2.5$ and $\alpha = -3$) and for various data-set lengths (100, 400, 700, 1000, 1300, 1600, 1900 and 2200 t.u.) we produce 2000 artificial light curves. In total we perform 40 simulations, each one consisting of 2000 artificial light curves. Then, for each PSD-index and light curve length we localise the position of the τ_{br} , as described in Section 3.1, and after forming its distribution we estimate its mean and its standard deviation³ (Fig. 5).

For all the PSD-slopes we can readily see that as the length of the data set increases from t t.u. to t' t.u. ($t' > t$), the SF-break gradually shifts to larger values. This can be understood for the case of a random walk process ($\alpha = -2$) where the mean distance after T steps is \sqrt{T} . The mean

distance in our case is the mean variance which is measured from the SF, hence for $t' > t$ the maximum variance and hence the position of the break is expected to shift towards larger time-scales.

The important point here is that the measured variance of a data set (and hence the maximum variance depicted by the break) does not reflect the true underlying variability properties of the source. As the data set increases, larger variations produced from the same variability process having the same PSD enter the data sets and this is mapped in the SF as a break even in the case of featureless PSD which are lacking characteristic time-scales.

In Fig. 5 we show the position of the SF-breaks, τ_{br} , induced by the different length of the data sets for various PSD-slopes. This diagram can be used only for continuously sampled light curves for which the underlying PSD is a pure power-law with the given indices. The existence of data-gaps in the data set can alter significantly this picture since sampling patterns play a major role in the form of the SF and hence the position of the break (Section 6). In the case that our data set has a PSD of a broken-power-law form, similar simulations should be performed in order to specify the position of the fake SF-break. Note that the physically interesting PSD-breaks will be mapped in the SF but in statistically unsound way with respect to their position and their uncertainties (Section 5.3).

Consider the case of the blazar PKS 2155-304 as presented by Zhang et al. (2002). The 1997 (~ 100 points) and 1999 (~ 200 points) data sets are the most continuous ones (figure 2 and figure 3 in Zhang et al. 2002) and they have featureless PSD with slopes $\alpha = -2$ and $\alpha = -3$ respectively (Table 3 in Zhang et al. 2002). The SF of the corresponding light curves (figure 6, middle right panel and bottom right panel respectively, in Zhang et al. 2002) exhibit breaks of the order of $\tau_{br} \approx 12.8$ ksec and $\tau_{br} \approx 61.7$ ksec respectively (or 13 t.u. and 61.7 t.u. respectively for SF-bins of 1 ksec). A SF-break for the 1997 data set is predicted (from Fig. 5) to lie at around 14.8 ± 1.4 t.u. and for the 1998 data set, at around 60 t.u. with a simple linear interpolated estimation. This outcome readily tells us that these breaks can originate from a variability process with no characteristic time-scale simply described by a featureless PSD of a power-law form.

4 THE STRUCTURE FUNCTION AND THE BLAZAR SHOT-NOISE MODEL

Particle acceleration in shocks (Kirk et al. 1998) is commonly invoked to explain the observed temporal and spectral variability of blazars. There may be a variety of intrinsic timescales, i.e. acceleration, escape and light crossing timescales, associated with a single emitting region. TAN01 simulate light curves on the assumption of certain shot parameters and, from the similarity between the simulated and observed SFs, estimate some temporal shot properties. Here we caution against this approach as, in general, the light curves derived from such shots do not exhibit the same PSD as those observed from the blazars. A physically correct model should reproduce both the SF and the PSD.

In TAN01 a number of simulations were carried out in order to associate the SF-breaks with the properties of the

³ The distributions are very close to Gaussian (see e.g. Fig. 2).

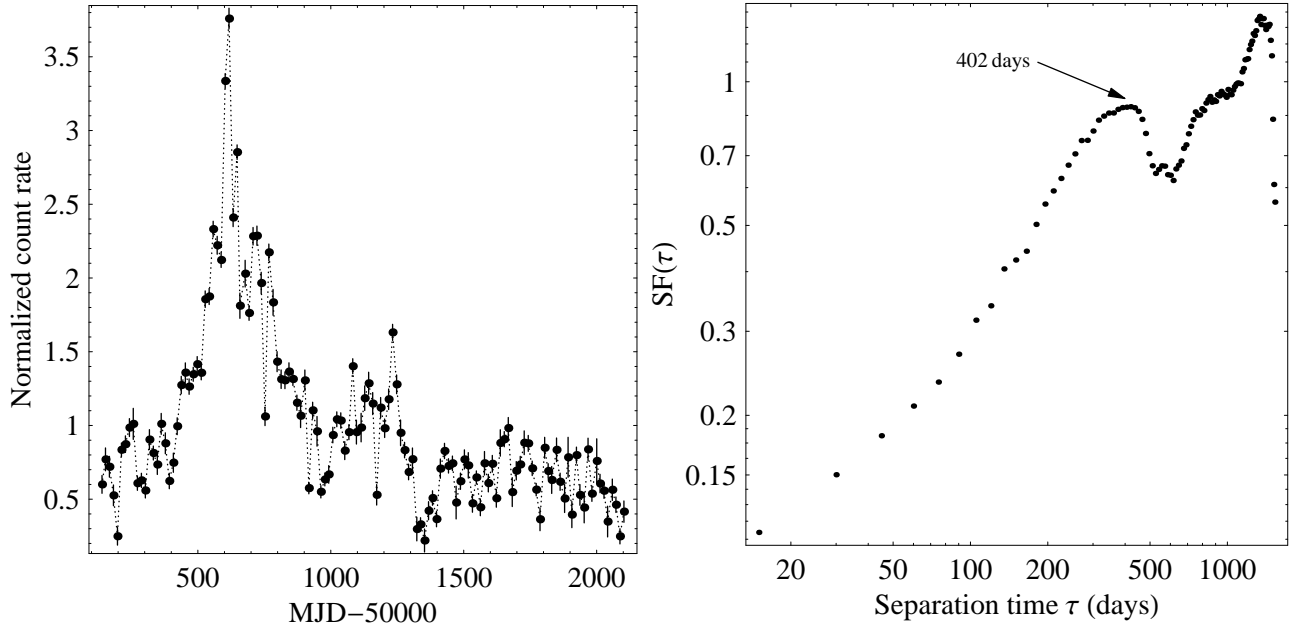


Figure 3. [Left panel] The long-term 2–10 keV *ASM* light curve of Mrk 501, between 50100 MJD–52100 MJD, in bins of 15-days. The dotted line is linear interpolation intended to guide the eye. [Right panel] The SF of the *ASM* light curve of Mrk 501, in bins of 15-days (in logarithmic scale). The SF-break occurs around 402 days.

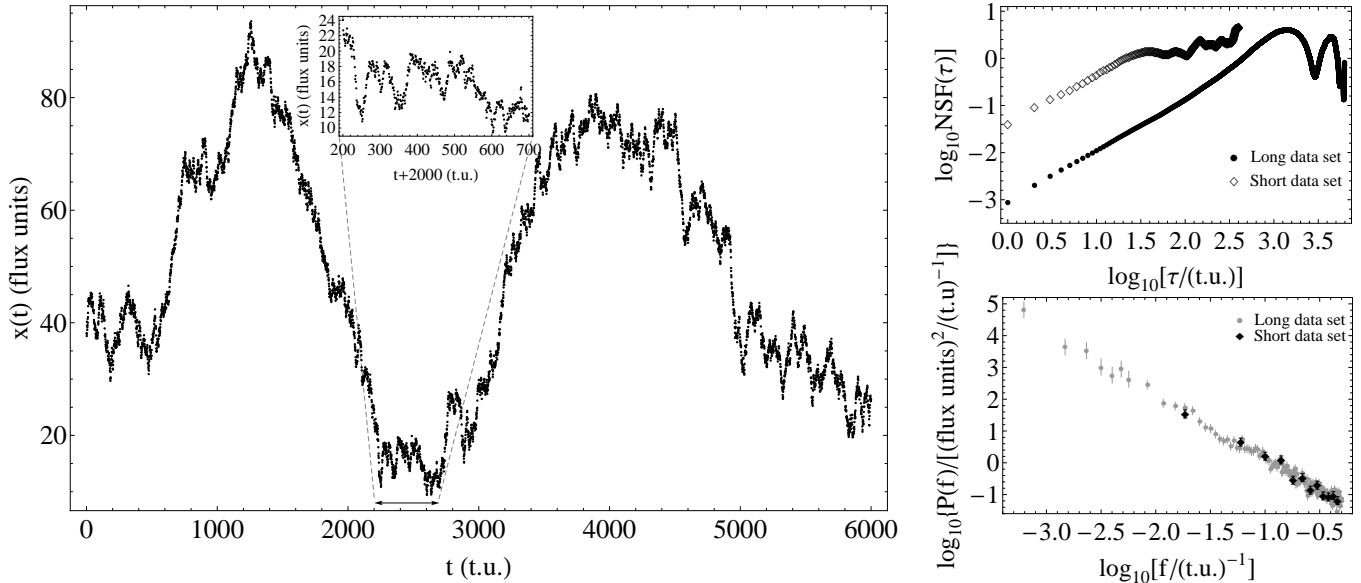


Figure 4. [Left panel] The overall simulated light curve 6000 t.u. long produced from an underlying PSD having a power-law form with index -2. The inlay shows a short segment of the light curve 500 t.u. long. [Top right panel] The NSF of the data in the inlay (empty diamonds) and the overall data set (black points). The break of the SF occurs at 40 and 1600 t.u. respectively. [Bottom right panel] The binned logarithmic periodograms of the short (filled diamonds) and the overall data sets (grey points). Both of them match the underlying PSD, being a featureless power-law of index -2.

observed flares. The light curve is regarded as a superposition of *triangular shots*⁴ with rise and decay time-scale t_r

and t_d respectively, occurring randomly at t_p following a

⁴ The last 40 years the commonly used shot-noise models consist of pulses with either a symmetric exponential rise and decay pro-

file or only an exponential decay profile (e.g. Lochner et al. 1991; Burderi et al. 1997).

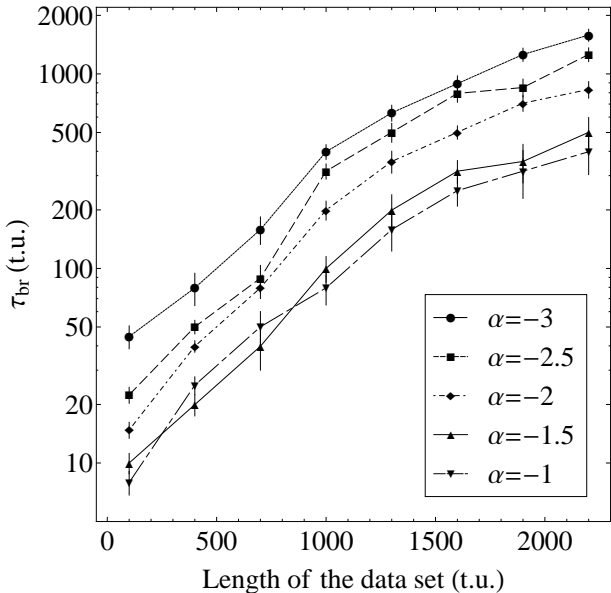


Figure 5. The position of the SF-break, τ_{br} (in logarithmic scale), as a function of the length of the data set assuming an underlying PSD of a power-law form of index α . Each data point represents the mean value and the standard deviation of the distributions of the SF-breaks yielding from an ensemble of 2000 simulations having a power-law PSD of index -1 (filled circles), -1.5 (filled squares), -2 (filled diamonds), -2.5 (filled up-pointing triangles) and -3 (filled down-pointing triangles). The dashed lines are linear interpolations intended to guide the eye.

Poisson distribution, with an intensity I_0 at t_p

$$\mathcal{I}_{tr,sh.}(t) = \begin{cases} \frac{I_0}{t_r} [t - (t_p - t_r)] & t_p - t_r < t \leq t_p \\ -\frac{I_0}{t_d} [t - (t_p - t_d)] & t_p < t < t_p + t_d \end{cases} \quad (2)$$

Then, the SF properties were studied with respect to the aforementioned time-scales and the main conclusion was that only the shortest time-scale between t_r and t_d determines the position of the break. Moreover, due to the fact that the *ASCA* data set of Mrk 501 consists mainly of symmetric flares, they connect the shortest time-scale derived from the SF to the light crossing time of the emission region.

The apparent similarities between the observed SFs and those coming from the shot-noise simulations, presented by TAN01, give an erroneous impression that such correspondences hold. Apart from the fact that breaks in the SF can occur in data sets with no characteristic time-scales (Section 3), we posit that the shot-noise model is not a physically realistic approach that can be used in order to associate the observed breaks of the SF to the smallest time-scales embedded in the data sets. In order for the shot-noise simulations to be realistic representations of the observed *ASCA* data set of Mrk 501, they must be able to reproduce the observed PSD as well as the SF.

We repeat the simulations of TAN01 by creating shot-noise light curves 2000 t.u. long and we calculate the corresponding PSD. Since the general shape of the latter can be understood qualitatively by the PSD form of the individ-

ual triangular shots, used for the construction of each light curve, it is useful to look at the general analytical form of the PSD for a single triangular shot as derived directly from the Fourier transform of the continuous functions (Appendix C). We should note here that since the simulated light curves are discretised, apart from their noisy form, due to the randomness of the occurrence times and of the intensities, we expect the PSD to flatten towards high frequencies due to aliasing effects (e.g. Papadakis & Lawrence 1993).

Initially TAN01 consider that the light curve consists of identical symmetric triangular shots i.e., $\tau = t_r = t_d = 10$ t.u. The PSD of such a process breaks at τ^{-1} and becomes very steep at higher frequencies with power-law of index $\alpha \approx -4$ superimposed on a wave with peaks separated by τ^{-1} (Fig. 6, panel a). Qualitatively we can justify this behaviour by considering the PSD of a single symmetric triangular shot which is proportional to $f^{-4} \sin^4(\pi\tau f)$ (equation (C3)). We expect that in the case of randomly occurring symmetric triangular shots, having different intensity but the same duration, the overall shape of their PSD should be similar to the sum of the PSDs of the individual symmetric triangular shots.

The second case considered by TAN01 is that of non-identical symmetric shots ($\tau = t_r = t_d$) whose time-scale τ varies randomly between $\tau_{min} = 10$ t.u. and $\tau_{max} = 100$ t.u. In this case the PSD is flat until τ_{max}^{-1} and then it breaks following a simple steep power-law of index $\alpha \approx -4$ (Fig. 6, panel b). The oscillatory behaviour, caused by the sine term of equation (C3), is smoothed out due to the random distribution of τ s coming from every symmetric shot.

The third and the last case considered by TAN01 is that of asymmetric identical shots with $\tau_r = 10$ t.u. and $\tau_d = 100$ t.u. This is actually a generalization of the first case. The PSD breaks at τ_d^{-1} , then it becomes steep with a power-law index of $\alpha \approx -2$ and after τ_r^{-1} beats separated by τ_r^{-1} are superimposed on the underlying power-law (Fig. 6, panel c). Qualitatively, based on the PSD of a single asymmetric shot (Fig. C1) we can readily distinguish very similar spectral features such as the position of the first break, the slope between τ_d^{-1} and τ_r^{-1} , and the separations in frequency of the beating frequencies.

Each of the aforementioned simulated PSDs differs significantly from the PSD derived from the *ASCA* observations i.e. a simple power-law with index $\alpha = -1.80 \pm 0.09$. Neither steep slopes nor sinusoidal features in oscillating or beating forms appear in the featureless PSD of Mrk 501. The fact that the PSD functions differ significantly between simulations and observations clarifies in an unambiguous way that the variability properties of Mrk 501 can not be described in a physically correct way from the shot-noise model. The latter can very well reproduce the various SF-features i.e. slopes and breaks observed in the actual SFs of the observed in blazars, but it can not reproduce the featureless PSD. This means that the shot-noise simulations are not representative of the true underlying variability properties of Mrk 501 and hence should not be used in order to associate the observed SF-breaks with characteristic physical time-scales.

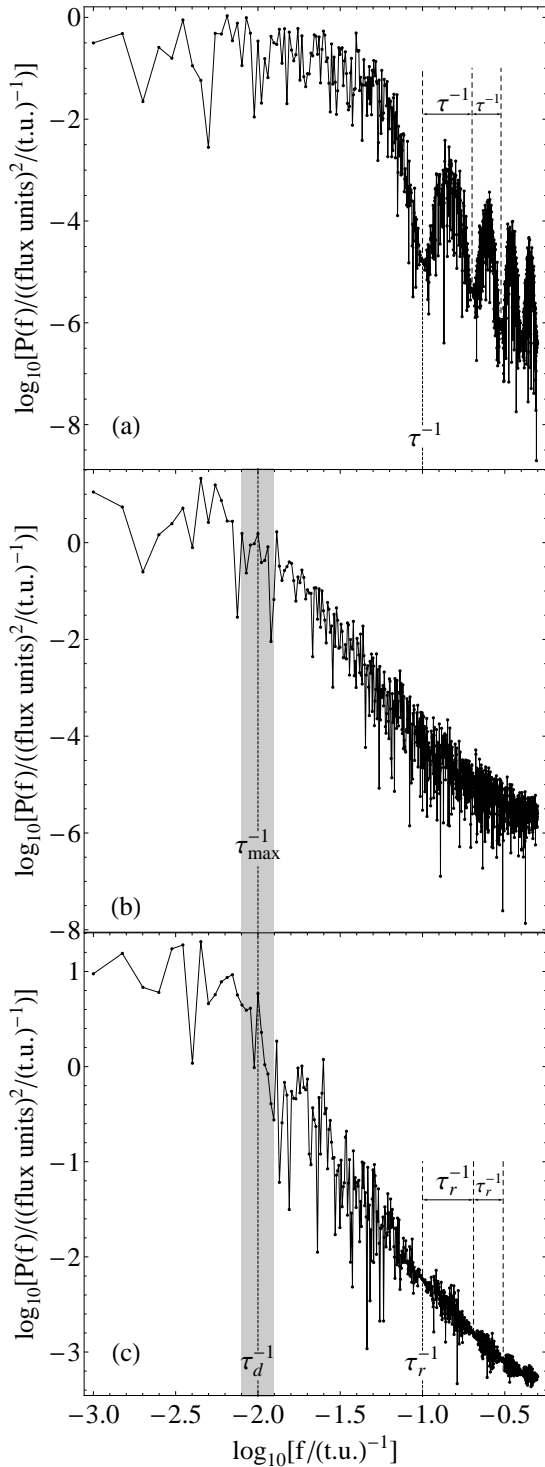


Figure 6. Example PSDs for the shot-noise model consisting of triangular shots for three cases of shots.

[Top panel (a)] Identical symmetric shots with rise and decay time, τ_r and τ_d respectively, equal to 10 t.u.

[Middle panel (b)] Symmetric shots with rise and decay time equal to τ ($\tau_r = \tau_d = \tau$), and τ distributed between $\tau_{\min} = 10$ t.u. and $\tau_{\max} = 100$ t.u.

[Bottom panel (c)] Asymmetric identical shots with rise and decay time $\tau_r = 10$ t.u. and $\tau_d = 100$ t.u., respectively.

5 FITTING MODELS TO THE STRUCTURE FUNCTION: PARAMETER AND ERROR ESTIMATION

5.1 Lack of Gaussianity and statistical independence

One of the major problems that affects the use of SF is that the various estimates, $SF(\tau_i)$, are not independent of each other. This problem severely affects the fitting routines e.g. least-squares, maximum likelihood, that are commonly used in the published blazar-SF-literature to derive the SF-breaks and the -slopes. These routines all require that the data points should be statistically independent so that the probability of obtaining the full ensemble, in our case $SF(\tau_i)$ for all τ_i , is equal to the product of the probabilities of obtaining the individual $SF(\tau_i)$ (e.g. Bevington & Robinson 1992).

That means that if we want to fit a broken-power-law of the form (e.g. Zhang et al. 2002)

$$SF_M(\tau; C, \tau_{\max}, \beta_1, \beta_2) = \begin{cases} C \left(\frac{\tau}{\tau_{\max}} \right)^{\beta_1} & \tau \leq \tau_{\max} \\ C \left(\frac{\tau}{\tau_{\max}} \right)^{\beta_2} & \tau > \tau_{\max} \end{cases} \quad (3)$$

to a data set of the form

$$\left\{ (\tau_1, SF(\tau_1)), (\tau_2, SF(\tau_2)), \dots, (\tau_N, SF(\tau_N)) \right\}$$

we have to maximize the probability of obtaining the ensemble of estimates $\{SF(\tau_1), SF(\tau_2), \dots, SF(\tau_N)\}$, known as the *joint probability*

$$P(C, \tau_{\max}, \beta_1, \beta_2) = \prod_{i=1}^N P_i(SF_M(\tau_i; C, \tau_{\max}, \beta_1, \beta_2)) \quad (4)$$

where $P_i(SF_M(\tau_i; C, \tau_{\max}, \beta_1, \beta_2))$ is the probability of obtaining the estimate $SF(\tau_i)$. Equation (4) is valid (i.e. the joint probability equals to the product of individual probabilities) *iff* the estimates $\{SF(\tau_1), SF(\tau_2), \dots, SF(\tau_N)\}$ are independent of each other.

Unfortunately the adjacent SF-estimates are far from independent. In Fig. 7 we show the *autocorrelation function* (ACF) of the first 25 logarithmic SF-estimates (out of 137) of Mrk 501 (Fig. 1) after detrending them by subtracting a fifth degree polynomial. Note here that this detrending is crucial since we want to check the linear relations between the SF-estimates (e.g. short-term dependence), and not long-term linear-dependence which is due to the general shape of the SF. In Fig. 7 we quantify the degree of linear dependence of the SF-estimates for the slope, the break and the plateau regions. It is impossible to have linearly correlated values which are independent⁵. We can readily see that up to $\tau = 0.88$ days (the range which is used in order to derive the break time-scale in Section 3.1) the degree of linear correlation is greater than 0.5.

The assumption of Gaussianity is another issue with

⁵ We should note that sometimes the absence of a linear correlation is confused with statistical independence but this is the case only when we are dealing with Gaussian distributions. Genuine statistical independence requires not only linearly uncorrelated values but also the absence of any functional relation between the variables.

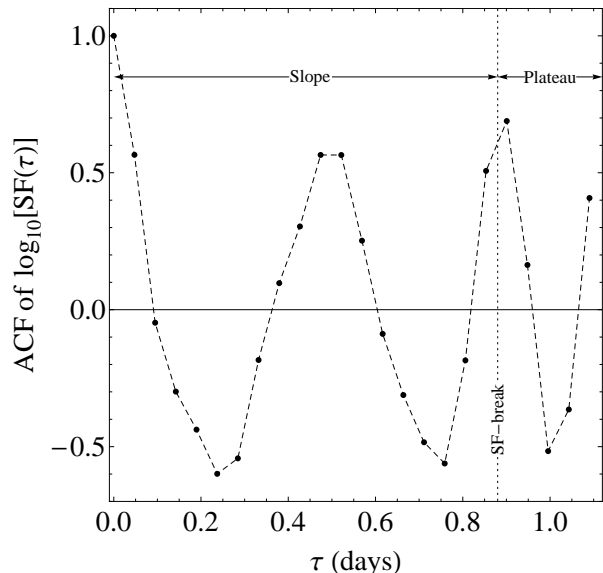


Figure 7. The ACF of the first 25 detrended logarithmic SF-estimates of Mrk 501 (from Fig. 1) showing the degree of linear dependence among them. The vertical dotted line shows the position of the SF-break at 0.88 days, and the arrows indicate the SF-slope and plateau regions. The dashed lines among the points of the ACF are linear interpolations intended to guide the eye.

the SF fitting procedures. Only if the distribution of the SF-estimates within each τ_i is Gaussian (with a mean value $SF(\tau_i)$ and standard deviation σ_i), is it valid to consider that

$$P_i(SF_M(\tau_i; C, \tau_{\max}, \beta_1, \beta_2)) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left[\frac{SF(\tau_i) - SF_M(\tau_i; C, \tau_{\max}, \beta_1, \beta_2)}{\sigma_i} \right]^2} \quad (5)$$

and estimate the product in the joint probability (equation (4)) (which implies independence) as

$$P(C, \tau_{\max}, \beta_1, \beta_2) = \prod_{i=1}^N \left(\frac{1}{\sigma_i \sqrt{2\pi}} \right) e^{-\frac{1}{2} \sum_{i=1}^N \left[\frac{SF(\tau_i) - SF_M(\tau_i; C, \tau_{\max}, \beta_1, \beta_2)}{\sigma_i} \right]^2} \quad (6)$$

According to the maximum likelihood method, in order to find the most probable set of parameters $(C, \tau_{\max}, \beta_1, \beta_2)$ we must maximize $P(C, \tau_{\max}, \beta_1, \beta_2)$ (equation (6)) or equivalently we must minimize the sum-argument χ^2 in the exponential of equation (6)

$$\chi^2 = \sum_{i=1}^N \left[\frac{SF(\tau_i) - SF_M(\tau_i; C, \tau_{\max}, \beta_1, \beta_2)}{\sigma_i} \right]^2 \quad (7)$$

Thus the quantity χ^2 automatically defines the goodness-of-fit which is actually a measure of the probability of obtaining the observed SF-estimates from a given set of parameters $(C, \tau_{\max}, \beta_1, \beta_2)$ but only when we are dealing with independent and Gaussian SF-estimates.

However, in reality we do not deal with Gaussian SF-estimates. By producing 2000 random data sets 500 t.u. long from a PSD of a simple power-law form having index -2, we form the distribution of the SFs for the first, the second, and the three-hundredth time bins respectively (Fig. 8, top panels). The histograms clearly have a non-Gaussian form and

remains non-Gaussian for all the time bins τ . That means that any fitting relation of the form of equation (3), dealing directly with the SF-estimates, does not provide reliable results since Gaussianity i.e. equation (5), is not valid.

Interestingly, at least for those time bins which are below the fake SF-break, occurring at $\tau_{br} \approx 50$ t.u. (Fig. 5), the logarithms of the SF-estimates follow a distribution which is closer to Gaussian (Fig. 8, bottom panels a and b). For $\tau > \tau_{br}$ even the logarithmic-estimates of the SF deviate significantly from Gaussianity (Fig. 8, bottom panel c) and of course for all the time bins τ the measurements continue to be statistically dependent on each other.

Kataoka et al. (2001) define the sum of square differences $\chi_{sim}^2 = \sum_k \{ \log_{10}[SF(\tau_k)] - \log_{10}[SF(\tau_k)] \}^2$, where the angle brackets indicate the arithmetic mean, in order to derive a statistical significance of the goodness of their SF fit, accounting in this way for the non-Gaussianity. They argue that χ_{sim}^2 is different from the traditional χ^2 (i.e. equation (7)) but the statistical meaning is the same. Since the SF measurements are not independent, a χ^2 function of the form $\sum(S - E)^2$, where S comes from simulations (i.e. from a given assumed underlying model) and E are the actual estimates, makes sense only if for every SF bin the distribution of the entries is Gaussian (yielding the $(S - E)$ from the exponent of the Gaussian, equation (5)) and if all the SF bins are independent of each other (yielding the sum from the joint probability, equation (6)).

Concerning the Gaussianity it would be more appropriate for Kataoka et al. (2001) to use the mean logarithm of the SF, $\langle \log_{10}[SF(\tau_k)] \rangle$ rather than the logarithm of the mean SF, $\log_{10}[SF(\tau_k)]$, in their pseudo chi-square estimate: $\chi_{sim, Gauss}^2 = \sum_k \{ \langle \log_{10}[SF(\tau_k)] \rangle - \log_{10}[SF(\tau_k)] \}^2$ where both quantities $\langle \log_{10}[SF(\tau_k)] \rangle$ and $\log_{10}[SF(\tau_k)]$ are distributed Gaussian within a t_k . However, the main problem of non-independence is not avoided and therefore the derived significances are not statistically meaningful.

5.2 Error underestimation

In most of the cases, when we deal with statistically independent variables under the assumption of Gaussianity, the most probable value (i.e. the one that minimizes the χ^2) and its standard deviation are enough to give a full picture of the distribution of the fit parameters.

There are several methods, as we are going to see, that are used in the literature claiming robustness and statistically meaningful errors for the SF fit parameters i.e. slopes and breaks. In this section we show through simulations that the lack of statistical independence in the SF-estimates yields very small estimates of uncertainty in the fitting parameters, which do not reflect the true underlying distributions of the fitting parameters. The essence of any correct fitting procedure is to yield a statistically meaningful description of these distributions.

5.2.1 Fitting the SF-slopes and -breaks

We use the previously generated simulations (2000 light curves, 500 t.u. long from a power-law PSD having an index -2), and for each light curve we calculate the logarithm of the SF-estimates. We attribute to every estimate a standard error based on the error of the sample mean for each

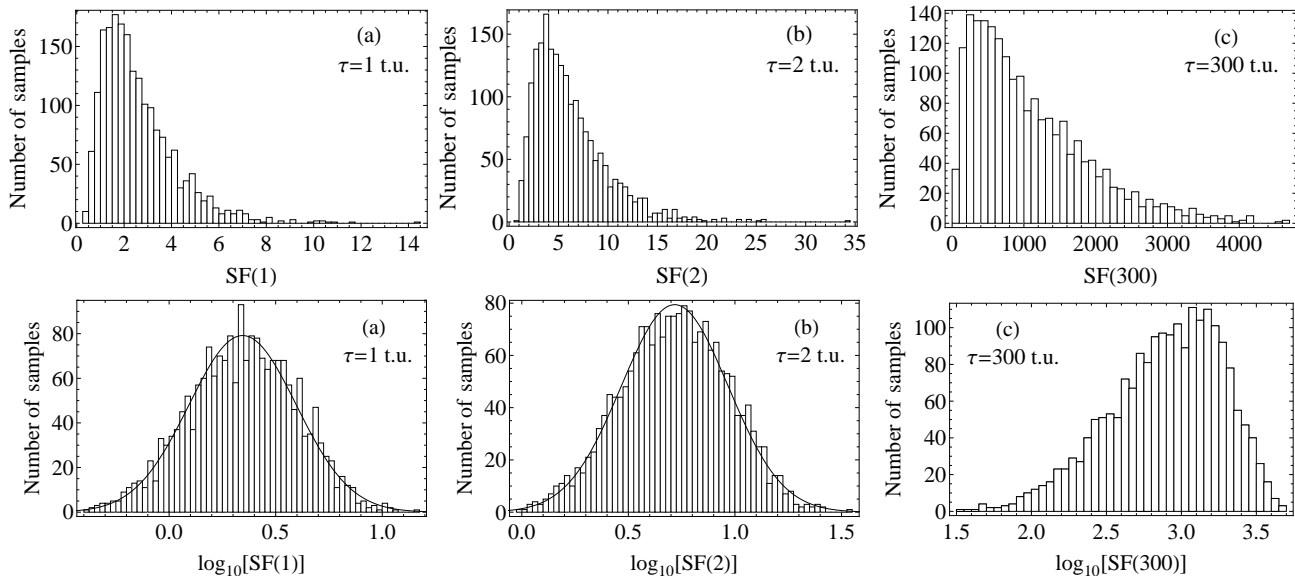


Figure 8. The distributions of the ensemble of 2000 SF-estimates, coming from 2000 random data sets 500 t.u. long having a PSD of a power-law form of index -2, for the first, second and three-hundredth time bin respectively.

[Top panels] The distributions of the normal SF-estimates (the histogram bins have a length of 1, 1 and 100 respectively) is not Gaussian for all the τ .

[Bottom panels] The distribution of the logarithmic SF-estimates (the histogram bins have a length of 0.025, 0.025 and 0.05 respectively) approximates adequately the Gaussian distribution only for $\tau \lesssim 50$ t.u.

time bin, which is one of the most usual methods (e.g. Collier & Peterson 2001; Zhang et al. 2002; Czerny et al. 2003). We then fit, using a least-squares fitting procedure, a simple linear model of the form

$$\log_{10}[\text{SF}(\tau; C, \beta)] = C + \beta \log_{10}(\tau) \quad (8)$$

where β represents the slope of the SF. Hence, for each light curve we estimate the value of the SF-slope β together with an error $\delta\beta$. The top panels of Fig. 9 show the distribution of β and $\delta\beta$. The errors derived from the fit, $\delta\beta$, are very small in comparison to the actual scatter of the measured β , depicting clearly the effect of non-independency. From the simulations we expect 68 per cent of the derived slopes to be within the interval 1.02 ± 0.13 but erroneously the $\delta\beta$ distribution has a mean of 0.006, dictating a range for the β within 1.020 ± 0.006 . The reason for the very small uncertainties in the fitted slopes, $\delta\beta$, is that the errors on the actual SF-estimates are very small themselves due to their non-independent nature. Every SF-estimate follows smoothly the increasing trend, defined from the adjacent points, and thus the errors derived directly from the SF-estimates in this way are always too small.

On the contrary the binned logarithmic periodogram-estimates together with their corresponding errors do provide us with statistically correct information about the behaviour of the slope. For each one of the previous simulated light curves we fit the binned logarithmic periodogram-estimates with a simple linear model and once again we derive the slope α and the error $\delta\alpha$. We can see from the bottom panels of Fig. 9, that 68 per cent of the α -estimates are expected to fall within -1.98 ± 0.18 , in accordance with the distribution of $\delta\alpha$ (having a mean of 0.17) coming directly from the fits to the binned logarithmic periodograms of the simulated light curves. A meaningful error analysis originat-

ing from a fitting procedure, is correct only when its predictive character (i.e. in our case for the PSD: 68 per cent of the measurements should be within the range -1.98 ± 0.18) coincides with the actual fluctuational outcome of the process (i.e. -1.98 ± 0.17). These simulations show us clearly that this is not the case for the SF errors, which are much smaller than the true spread of the SF-estimates for the same variability process.

Occasionally authors attempt to take account some of the SF problems. In Agudo et al. (2006), for the case of the intra-day variable blazar S50716+71 observed at 86 GHz, the authors use the interpolated version of the SF which takes into account only the errors caused by the interpolation (Quirrenbach et al. 2000). By repeating this procedure for our simulations we get a mean error in the slope of 0.011 which is very small compared to the expected 0.13. Another weakness of this method is that several points in the SF have zero uncertainty (i.e. the intersection point of the forward and backward interpolations, see figure 4 in Agudo et al. 2006 and figures 24–40 in Quirrenbach et al. 2000) therefore they do not really have any statistical meaning.

In Fuhrmann et al. (2008), again for the case of S50716+71 but observed at radio frequencies, the authors make use of the term “variability timescale” which they relate directly to the SF-break. To estimate its uncertainty they use the saturation level p_0 and the two fitted parameters of the SF function α , β , $\text{SF}(\tau) = a\tau^\beta$, (see equations (B.2) and (B.3) in the appendix of Fuhrmann et al. 2008). We use the simulations of Section 3.1 in order to see how well we can specify the position of these spurious SF-breaks by applying the same methodology, despite the fact that the simulated light curves do not contain any real characteristic time-scale. The distribution of the spurious break time-scales is shown in the left panel of Fig. 2 (grey area), having

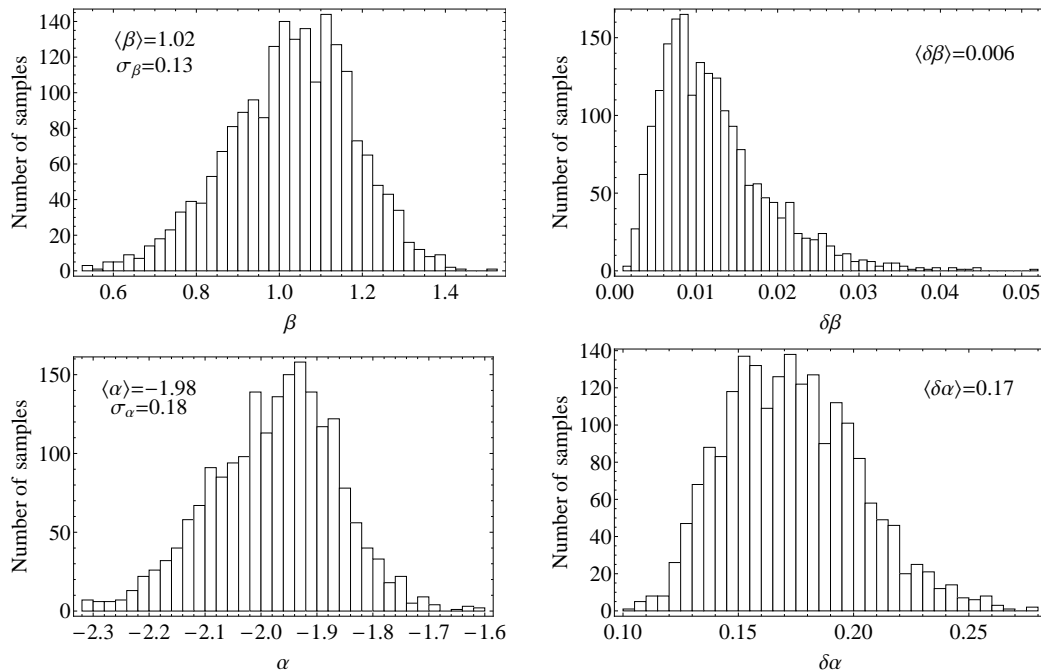


Figure 9. The distributions of SF- and PSD-slopes estimated from an ensemble of 2000 artificial light curves having an underlying PSD of a power-law form with slope -2.

[Top panels] (Left) The distribution of the SF slopes β has a mean of 1.02 and a standard deviation 0.13 (the histogram bins have a length of 0.025). (Right) The distribution of the errors coming from the fit $\delta\beta$, having a mean of 0.006 (the histogram bins have a length of 0.001).

[Bottom panels] (Left) The distribution of the PSD slopes α has a mean of -1.98 and a standard deviation 0.18 (the histogram bins have a length of 0.02). (Right) The distribution of the errors coming from the fit, $\delta\alpha$, having a mean of 0.17 (the histogram bins have a length of 0.005).

a mean value of 0.92 days and a standard deviation of 0.09 days. By applying the method of Fuhrmann et al. (2008) in the simulations of Section 3.1 we find a mean error for the breaks of 0.006 days which differs significantly from the expected 0.09 days. Obviously with such small (and incorrect) uncertainties every feature in the SF can be considered as a significant time signature of a special source property, as it seen in the right panel of figure 7 in Fuhrmann et al. (2008) where the authors find two “variability time-scales” embedded in their data set.

Each of the above cases, show us that timescales derived from the commonly-used methods to interpret and fit the SFs in the blazar literature, should be treated with great caution. Due to the lack of statistical independence, the errors on the fitting parameters tend to be always very small underestimating their true statistical scatter.

5.3 Resolving physically meaningful time-scales with the SF.

True characteristic time-scales embedded in the blazar light curves, should be mapped in the SF as well as in the PSD. In the limit of a single PSD slope in the range -1 to -3, covering a large frequency range and for a long light curve, there is a direct theoretical relation between the PSD and SF slopes (see Appendix B). However for the more common broken power-law PSD and with typical astronomical observing constraints, the relationships are less clear and thus, any attempt to relate SF-breaks or -slopes to PSD-breaks

or -slopes by one-to-one relationships can lead to misconceptions about the true underlying variability process.

We produce 2000 artificial light curves 500 t.u. long, having a PSD of a broken-power-law shape with a break at $f_{br} = 0.1 \text{ (t.u.)}^{-1}$ and with low and high frequency slopes of $\alpha_1 = -1$ and $\alpha_h = -2.5$ respectively. Then, for each light curve we estimate the SF and we produce an interpolated version of it in order to localize any feature around 10 t.u. in the form of either a local extremum or inflection point. We see a wide variety, and sometimes even an absence, of features around 10 t.u., something which places automatically the SF in the category of weak probing methods (Fig. 11). In this particular case we ensure that any feature around 10 t.u. is induced by the existence of the characteristic time-scale since we know *a priori* the expected position of it and hence no spurious break is expected at that time-scale. Of course this sort of interesting temporal signature has nothing to do with the previously mentioned (Section 3) physically uninteresting SF-break. As we can see from Fig. 11 these meaningless breaks, defined by the onset of a plateau on longer time-scales, are also present in our simulated light curves and their positions are simply deduced, as we showed in Section 3, from the length of the data set and the underlying PSD of the variability process. The distribution of the physically interesting SF-features, which are expected around 10 t.u., is shown in Fig. 10 having a mean value of 9.86 t.u. and a standard deviation of 1.63 t.u. coming from a total of 1789 SFs that exhibit features.

In order to resolve these physically interesting breaks,

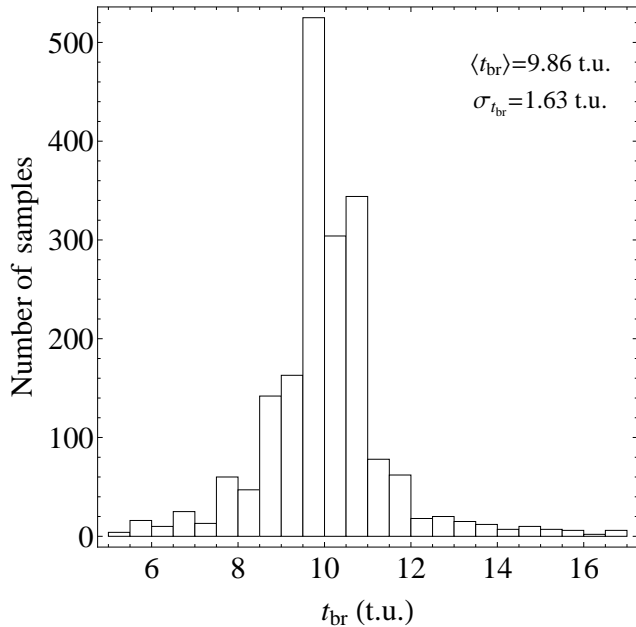


Figure 10. The distribution of SF-features (i.e. local extremum or inflection point) around 10 t.u., coming from 2000 artificial light curves 500 t.u. long, having a PSD of a broken-power-law shape with a break at $f_{br} = 0.1 \text{ (t.u.)}^{-1}$ and with low and high frequency slopes of $\alpha_1 = -1$ and $\alpha_h = -2.5$ respectively. The distribution has a mean value of 9.86 and a standard deviation of 1.63.

we employ the commonly used model of a broken-power-law, similar to the form of equation (3), but in its logarithmic version in order to ensure Gaussianity at least for the SF-estimates of the slope (Section 5.1).

$$\log_{10}[SF(\tau)] = \begin{cases} C + \beta_1 \log_{10}(\tau) & \tau \leq \tau_{br} \\ C + (\beta_1 - \beta_2) \log_{10}(\tau_{br}) + \beta_2 \log_{10}(\tau) & \tau \geq \tau_{br} \end{cases} \quad (9)$$

Fitting equation (9) to the simulated data typically results in the minimisation routine finding the physically uninteresting SF-breaks, t_{br} , occurring on long time-scales (≈ 100 t.u.) since this is the most prominent break-feature in the SF. Fortunately, and only for the case of the simulations, we know in advance the position of the expected SF feature we constrain our SF fits to the first 50 t.u., corresponding only to the increasing part of the SF, and we estimate the standard errors based on the error of the sample mean as described in Section 5.2. Moreover we constrain the range of τ_{br} between 2–20 t.u. in order to force the fitted τ_{br} parameter to correspond to the true PSD-break (Fig. 11, solid line).

The distribution of τ_{br} from the 2000 simulations, has a mean value of 5.31 t.u. and a standard deviation of 1.46 t.u. The shift of the most probable value from 9.86 t.u. to 5.31 t.u. is clearly a systematic problem of that method which depends on the form of the fitted model (equation (9)). Nevertheless, the scatter of these breaks, 1.46 t.u., is very close to the expected one, 1.63 t.u. Once again, the errors on τ_{br} derived directly from the fitting procedure are extremely small, having a mean value of 0.40 t.u. The broken-power-law model is not very representative of the actual shape of the SF which is continuously curved.

In order to model the physically interesting SF features we have to perform a much more detailed fitting even in this simple case tested here where we know in advance their approximate position. For this reason we employ a smoothly “double-bending” power-law (section 4.2 in McHardy et al. 2004) of the form

$$SF(\tau) = C\tau^{\kappa_1} \left[1 + \left(\frac{\tau}{\tau_{br1}} \right)^{\kappa_2 + \kappa_1} \right]^{-1} \times \left[1 + \left(\frac{\tau}{\tau_{br2}} \right)^{\kappa_3 - \kappa_2} \right] \quad (10)$$

with breaks at τ_{br1} and τ_{br2} and power-law indices κ_1 for $\tau < \tau_{br1}$, κ_2 for $\tau_{br1} < \tau < \tau_{br2}$ and κ_3 for $\tau > \tau_{br2}$. Again, we constrain the fitted break time-scales to be between 2–20 t.u. (Fig. 11, dashed line).

For the 2000 simulated light curves the distribution of the fitted position of the first break, τ_{br1} , has a mean value of 6.13 t.u. and a standard deviation of 1.32 t.u., and for the second break 15.03 t.u. and 1.45 t.u. respectively. Neither of these breaks depicts the actual position of the characteristic time-scale (10 t.u.). Similarly to the simple broken-power-law fitting model, the errors derived directly from the fit are very small, having a mean value of 0.35 and 0.39 for the two breaks respectively.

Despite the fact that breaks in the PSD, representing physically interesting variability time-scales, are mapped to the SF in the form of local maxima or inflection points, it is dangerous to make any statistical statement about them since

- (i) it is very difficult to depict the breaks robustly based on a fitting procedure due to the awkward shape of the SF; even if we know their position, based on the PSD, in which case we do not need to use the SF in the first place.
- (ii) the estimated errors in the fitted parameters (derived from a $\Delta\chi^2 = 1$, under the assumption of Gaussianity) are always very small with respect to the true scatter of the parameters due to the statistical dependence of the SF-estimates; even if we find a well behaved function able to depict the break based on single or multiple fitted parameters.

In reality, we deal with astronomical data sets for which we want to reveal the possible existence of characteristic time-scales in them. In the case that these time signatures do exist, we do not know in advance their position and how they are mapped on the SF. Moreover, since the SF-estimates are affected by the statistical properties of the observed light curves (i.e. mean, variance and the length) and not from the actual statistical properties of the underlying variability process (as is the PSD), even in the absence of characteristic time-scales we expect to see SF-artefact-features similar to the ones induced by real characteristic time-scales.

6 THE STRUCTURE FUNCTION AND DATA-GAPS

The SF method is considered by several researchers (e.g. Takahashi et al. 2000; Collier & Peterson 2001; Kataoka et al. 2001, 2002; Zhang et al. 2002) to be the ideal method of studying the time properties of gappy

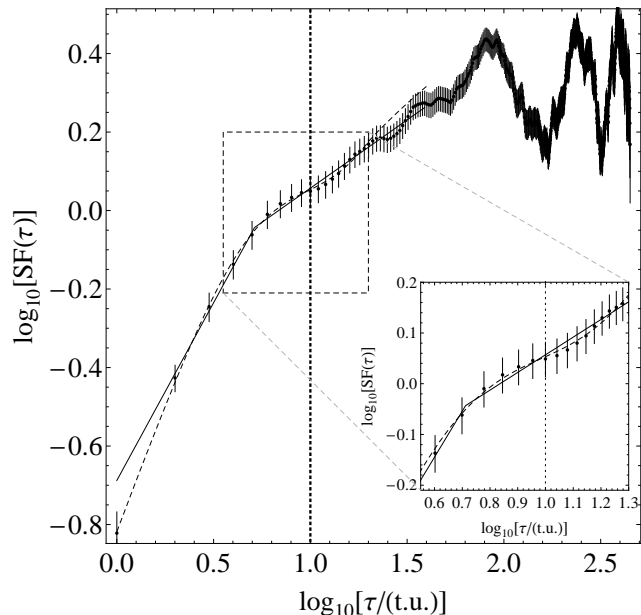


Figure 11. The points represent the SF-estimates for a randomly selected light curve, drawn from the ensemble of 2000 data sets 500 t.u. long, having a PSD of a broken-power-law shape with $f_{br} = 0.1 \text{ (t.u.)}^{-1}$, $\alpha_1 = -1$ and $\alpha_h = -2.5$ respectively. The thick-dotted vertical line indicates the position of the expected characteristic time-scale of 10 t.u. The specific SF has an inflection point at 10.32 t.u. A physically meaningless break appears at $\tau \sim 100$ t.u.

The solid line represents the SF model-fit (equation (9)) employed to depict τ_{br} which is constrained between 2–20 t.u., and yields $\tau_{br} = 5.15 \pm 0.43$ t.u. The dashed line represents the SF model-fit (equation (10)) employed to depict τ_{br} which is constrained between 2–20 t.u., yielding $\tau_{br1} = 5.97 \pm 0.38$ t.u. and $\tau_{br2} = 14.92 \pm 0.37$ t.u.

The inlay shows the behaviour of the two fitted models around the expected characteristic time-scale of 10.32 t.u.

data sets as it is believed to be less distorted by gaps than frequency-domain methods. Here, we use simulations to determine whether this confidence in the robustness to data gaps of the SF method is really justified.

Initially we produce a single artificial light curve 2000 t.u. long from a featureless PSD with a power-law shape with index -1.5 and we compute its SF (Fig. 12, panels 1a and 1b, respectively). We again note the spurious break at 315 t.u. (as expected from Fig. 5) which does not correspond to any real time property of the underlying variability process.

We then add to the simulated light curve three different types of data-gaps and we estimate the SF (Fig. 12, left panels). The first sampling scheme represents the case of *almost periodic data-gaps* where 57 per cent of the data are missing in almost periodic epochs, separated by 50–100 time units (Fig. 12, panels 1b). The second sampling scheme consists of *dense and sparse sampling* where data are equally spaced for small time periods, and more sparsely sampled for the remaining ones, yielding in total a data set having 83 per cent less data (Fig. 12, panels 1c). The last case that we consider is that of purely sparsely sampled data where 92 per cent of the data are missing (Fig. 12, panels 1d).

The effects of the abovementioned data-gaps on the SF are shown in the right panels of Fig. 12. To take into account

the uncertainties that are introduced by these gaps for the ensemble of our simulated light curves, we use the *bootstrap method*, discussed by Czerny et al. (2003). From each gappy light curve we select randomly, allowing repetition, 1000 subsamples of 2000 t.u. long, and we drop the multiple entries. Then we estimate the SF and from the ensemble of 1000 SFs we calculate for every time bin τ a standard deviation, including additionally the standard error in the sample mean coming from the gappy light curve.

We should emphasise that here we study the differences between the SFs coming from the gappy data sets (empty diamonds, right panels of Fig. 12) and those coming from the uninterrupted data set (grey points connected with the grey line, right panels of Fig. 12 which are also displayed in panel (1b) of Fig. 12). That does not mean that the uninterrupted SF is necessarily the “true”, “original”, or “parent” SF of the underlying variability process. This is only one SF, chosen randomly from an ensemble of light curves having a featureless PSD of a power-law form with index -1.5. Our intention is to check whether the observed SF is affected by the data-gaps since this is the method that several authors (e.g. Takahashi et al. 2000; Collier & Peterson 2001; Kataoka et al. 2001, 2002; Zhang et al. 2002) use to derive physically interesting results. A direct visual comparison of the SFs (Fig. 12) reveals that actually gaps do affect the shape of the SF since they introduce wiggles and bends. The latter can create the wrong impression that the underlying variability process is described by an underlying PSD of a broken-power-law form (as in Section 5.3).

At this point it is interesting to check whether or not the SF-errors estimated from the bootstrap method depict correctly the deviations between the SFs of continuously sampled and gappy light curves, SF_{conti} and SF_{gappy} respectively. We produce 2000 artificial light curves 2000 t.u. long, having a power-law PSD of slope -1.5, and we estimate for each one of them the SF_{conti} . Then, we apply to the light curves the dense and sparse sampling (example shown in Fig.12, panel 3a) and we estimate for each one of them the SF_{gappy} as well as an error for each time bin, $err_{boot,i}(\tau)$ (i.e. the standard deviation of the gappy-SF of the i^{th} simulated light curve at the time bin τ , $SF_{gappy}^i(\tau)$) as it is derived from the abovementioned bootstrap method. For each SF-bin we derive the following quantity

$$\left\langle \frac{|\Delta \log_{10}[SF(\tau)]|}{err_{boot,i}(\tau)} \right\rangle = \frac{1}{2000} \sum_{i=1}^{2000} \frac{|\log_{10}[SF_{gappy}^i(\tau)] - \log_{10}[SF_{conti}^i(\tau)]|}{err_{boot,i}(\tau)} \quad (11)$$

As we can see from Fig.13 the values of $\left\langle |\Delta \log_{10}[SF(\tau)]|err_{boot,i}^{-1} \right\rangle$ differ significantly from unity, having a mean value of 2.79 ± 0.91 . That means that the bootstrap method does not yield statistically meaningful errors that reflect the true deviations between the continuous and the gappy SFs. In other words, data gaps introduce *systematic deviations* in the structure function which depend on the light curve realization and cannot be accounted for without extensive simulations.

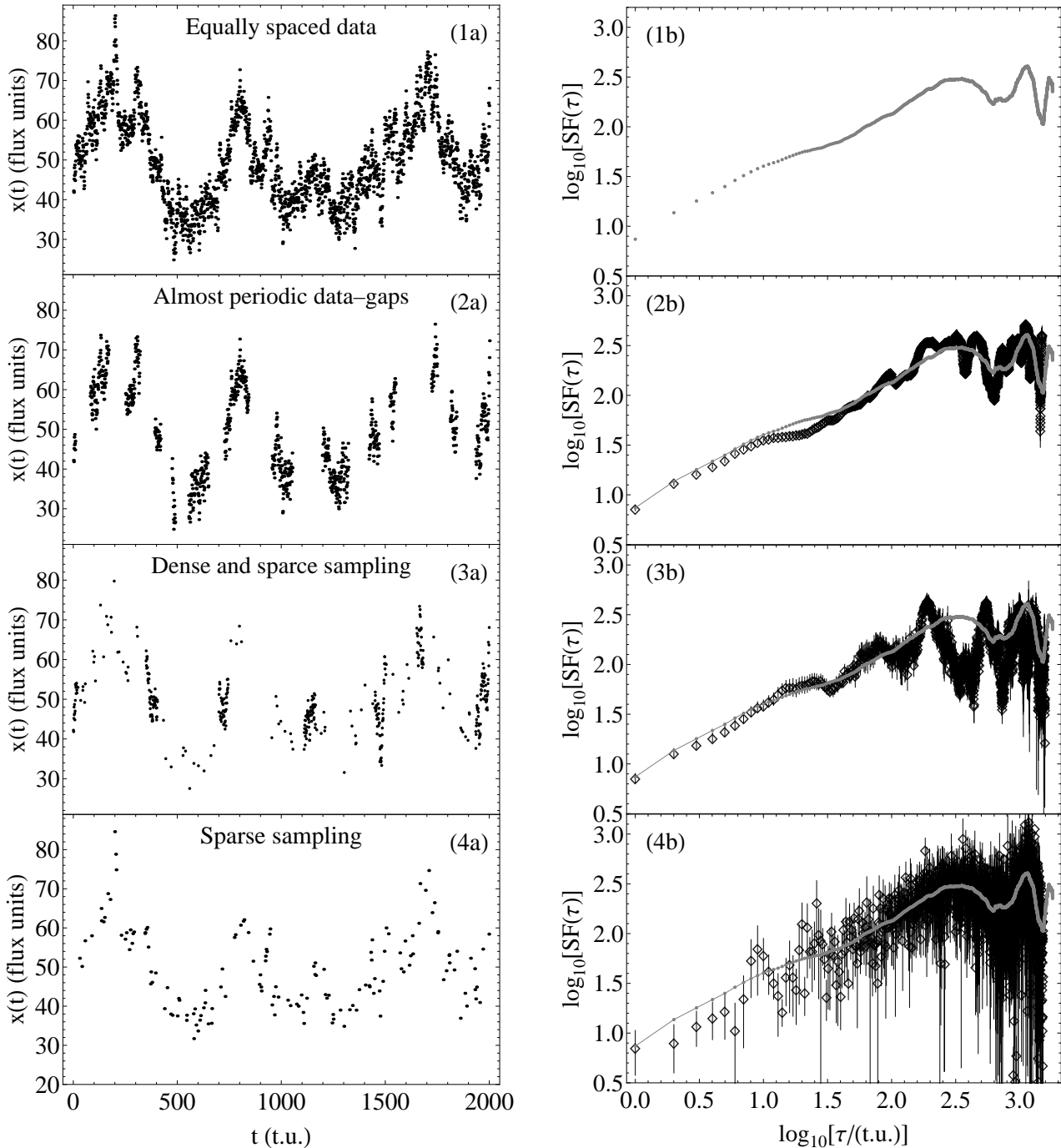


Figure 12. The effects of data-gaps in the SF of a simulated light curve (panel 1a) 2000 t.u. long, having a PSD of a power-law with index -1.5 . The left panels show the gappy patterns. The right panels show the corresponding gappy SF-estimates (black points) together with the continuous SF-estimates (grey points connected with grey line), as they are shown in panel 1b.

7 CONCLUSIONS

We have performed an extensive series of simulations in order to test the properties of the classical “running variance” SF-method. Under this fully controlled environment our studies have focused on the apparent SF-breaks, the correspondence of these breaks to the time-scales involved in the shot-noise model, the statistical behaviour of the SF with respect to the commonly used fitting procedures and

its response to data-gaps. Our results can be summarized clearly as follows:

- Strong SF-breaks frequently occur in data sets lacking any sort of characteristic time-scales. The position of these physically uninteresting breaks depends on the length of the observations and the shape of the underlying PSD.
- SFs derived from blazar observations resemble those coming from the shot-noise model i.e. superposition of triangular shots. However in the frequency domain the same

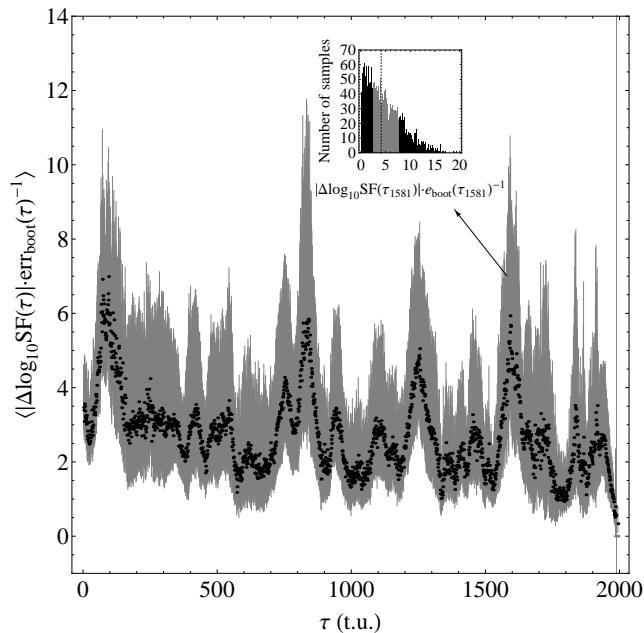


Figure 13. The mean difference between gappy and continuous SFs coming from an ensemble of 2000 artificial light curves, having the same underlying PSD of a power-law form with index -1.5. The gappy sampling scheme, applied to the continuous light curves, is shown in panel (3a) of Fig. 12. The errors $\text{err}_{\text{boot}}(\tau)$ are estimated for each light curve based on the bootstrap method (described in the text). The inset depicts the distribution of the points in the $\tau = 1581$ time-bin having a mean value of 4.16 and the standard deviation (i.e 34.1% of the entries) above and below the mean is shown in the grey region.

model does not show the same behaviour as the observations and thus is not physically realistic.

- Non-independence and non-Gaussianity mean that it is not possible to derive a meaningful goodness-of-fit from normal fitting procedures. We see, for example, that the derived uncertainties on the SF quantities i.e. positions of breaks and slopes, are always much less than the actual scatter of these variables during multiple realizations of the same variability process.

- Data-gaps affect severely the SF-estimates in an unpredictable way, introducing systematic deviations. The bootstrap method can not yield statistically meaningful errors depicting the deviations between the gappy and the continuous SFs.

We finally comment that for blazar variability present implementations merely determine the shape of the SF for the one realisation under consideration. That one realisation is, of course, just one of many possible realisations and so does not define precisely the properties of the true underlying variability process, which is really what we want to know. Current PSD analysis methods (e.g. Uttley et al. 2002) generally do, however, try to determine the shape of the underlying PSD by means of simulation-based modelling that take into account aliasing and other sampling effects. We encourage the blazar community to explore the use of these techniques or develop similar methodologies in the SF-field.

ACKNOWLEDGMENTS

DE and IMM acknowledge the Science and Technology Facilities Council (STFC) for support under grant ST/G003084/1. PU is supported by an STFC Advanced Fellowship. This research has made use of NASA's Astrophysics Data System Bibliographic Services.

REFERENCES

- Agudo I., et al., 2006, *A&A*, 456, 117
 Aharonian F., et al., 2003, *A&A*, 410, 813
 —, 2005a, *A&A*, 437, 95
 —, 2005b, *A&A*, 442, 895
 Aller M. F., Aller H. D., Hughes P. A., Latimer G. E., 1999, *ApJ*, 512, 601
 Bendat J. S., Piersol A. G., 1986, *Random Data: Analysis and Measurement Procedures*. New York: John Wiley & Sons, —c1986, 2nd ed.
 Bevington P. R., Robinson D. K., 1992, *Data reduction and error analysis for the physical sciences*. New York: McGraw-Hill, —c1992, 2nd ed.
 Blandford R. D., Kundic T., 1997, in *The Extragalactic Distance Scale*, Livio M., Donahue M., Panagia N., eds., pp. 60–75
 Bregman J. N., et al., 1988, *ApJ*, 331, 746
 Brinkmann W., Gliozzi M., Urry C. M., Maraschi L., Samburina R., 2000, *A&A*, 362, 105
 Brinkmann W., et al., 2001, *A&A*, 365, L162
 Burderi L., Robba N. R., La Barbera N., Guainazzi M., 1997, *ApJ*, 481, 943
 Catanese M., et al., 1997, *ApJ*, 487, L143+
 Coles W. A., Frehlich R. G., 1982, *Journal of the Optical Society of America (1917-1983)*, 72, 1042
 Collier S., Peterson B. M., 2001, *ApJ*, 555, 775
 Cordes J. M., Downs G. S., 1985, *ApJS*, 59, 343
 Czerny B., Doroshenko V. T., Nikolajuk M., Schwarzenberg-Czerny A., Loska Z., Madejski G., 2003, *MNRAS*, 342, 1222
 Done C., Madejski G. M., Mushotzky R. F., Turner T. J., Koyama K., Kunieda H., 1992, *ApJ*, 400, 138
 Fiedler R. L., et al., 1987, *ApJS*, 65, 319
 Fuhrmann L., et al., 2008, *A&A*, 490, 1019
 Gliozzi M., Brinkmann W., O'Brien P. T., Reeves J. N., Pounds K. A., Trifoglio M., Gianotti F., 2001, *A&A*, 365, L128
 Heeschen D. S., Krichbaum T., Schalinski C. J., Witzel A., 1987, *AJ*, 94, 1493
 Heidt J., Wagner S. J., 1996, *A&A*, 305, 42
 Hjellming R. M., Narayan R., 1986, *ApJ*, 310, 768
 Hufnagel B. R., Bregman J. N., 1992, *ApJ*, 386, 473
 Hughes P. A., Aller H. D., Aller M. F., 1992, *ApJ*, 396, 469
 Iyomoto N., Makishima K., 2001, *MNRAS*, 321, 767
 Kataoka J., et al., 2001, *ApJ*, 560, 659
 Kataoka J., Tanihata C., Kawai N., Takahara F., Takahashi T., Edwards P. G., Makino F., 2002, *MNRAS*, 336, 932
 Kirk J. G., Rieger F. M., Mastichiadis A., 1998, *A&A*, 333, 452
 Kolmogorov A. N., 1941a, *Doklady Akad. Nauk CCCP*, 30
 —, 1941b, *Doklady Akad. Nauk CCCP*, 32
 Lainela M., Valtaoja E., 1993, *ApJ*, 416, 485

Lewis G. F., Irwin M. J., 1996, MNRAS, 283, 225
 Lochner J. C., Swank J. H., Szymkowiak A. E., 1991, ApJ, 376, 295
 Maraschi L., et al., 1999, ApJ, 526, L81
 Markowitz A., et al., 2003, ApJ, 593, 96
 McHardy I. M., Arévalo P., Uttley P., Papadakis I. E., Summons D. P., Brinkmann W., Page M. J., 2007, MNRAS, 382, 985
 McHardy I. M., Gunn K. F., Uttley P., Goad M. R., 2005, MNRAS, 359, 1469
 McHardy I. M., Koerding E., Knigge C., Uttley P., Fender R. P., 2006, Nature, 444, 730
 McHardy I. M., Papadakis I. E., Uttley P., Page M. J., Mason K. O., 2004, MNRAS, 348, 783
 Neugebauer G., Soifer B. T., Matthews K., Elias J. H., 1989, AJ, 97, 957
 Papadakis I. E., Lawrence A., 1993, MNRAS, 261, 612
 Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical recipes in FORTRAN. The art of scientific computing. Cambridge: University Press, —c1992, 2nd ed.
 Priestley M. B., 1981, Spectral Analysis and Time Series: Probability and Mathematical Statistics, Vol. 1-2. London: Academic Press, —c1981
 Prokhorov A. M., Bunkin F. V., Gochelashvily K. S., Shishov V. I., 1975, Proceedings of the IEEE, 63, 790
 Quirrenbach A., et al., 2000, A&AS, 141, 221
 —, 1992, A&A, 258, 279
 Rickett B. J., Coles W. A., Bourgois G., 1984, A&A, 134, 390
 Romero G. E., Combi J. A., Benaglia P., Azcarate I. N., Cersosimo J. C., Wilkes L. M., 1997, A&A, 326, 77
 Rutman J., 1978, Proceedings of the IEEE, 66, 1048
 Scargle J. D., 1989, ApJ, 343, 874
 Sikora M., Błażejowski M., Begelman M. C., Moderski R., 2001, ApJ, 554, 1
 Simonetti J. H., Cordes J. M., Heeschen D. S., 1985, ApJ, 296, 46
 Smith A. G., Nair A. D., Leacock R. J., Clements S. D., 1993, AJ, 105, 437
 Spada M., Ghisellini G., Lazzati D., Celotti A., 2001, MNRAS, 325, 1559
 Stalin C. S., Gupta A. C., Gopal-Krishna, Wiita P. J., Sagar R., 2005, MNRAS, 356, 607
 Subba Rao T., Priestley M. B., Lessi O., 1997, Applications of time series analysis in astronomy and meteorology. London: Chapman and Hall, —c1997, 1st ed.
 Takahashi T., et al., 2000, ApJ, 542, L105
 Tanihata C., Takahashi T., Kataoka J., Madejski G. M., 2003, ApJ, 584, 153
 Tanihata C., Urry C. M., Takahashi T., Kataoka J., Wagner S. J., Madejski G. M., Tashiro M., Kouda M., 2001, ApJ, 563, 569
 Teräsraanta H., Wiren S., Koivisto P., Saarinen V., Hovatta T., 2005, A&A, 440, 409
 Timmer J., Koenig M., 1995, A&A, 300, 707
 Tosti G., et al., 1998, A&A, 339, 41
 Uttley P., McHardy I. M., 2005, MNRAS, 363, 586
 Uttley P., McHardy I. M., Papadakis I. E., 2002, MNRAS, 332, 231
 Vaughan S., 2005, A&A, 431, 391
 Vaughan S., Edelson R., Warwick R. S., Uttley P., 2003,

MNRAS, 345, 1271
 Yu C. X., Gilmore M., Peebles W. A., Rhodes T. L., 2003, Physics of Plasmas, 10, 2772
 Zhang Y. H., et al., 2002, ApJ, 572, 762

APPENDIX A: FUNCTIONAL RELATION BETWEEN THE STRUCTURE FUNCTION AND THE AUTOCORRELATION FUNCTION

Consider N observations of a stationary real-valued time process $x(t)$: $\{x(t_1), x(t_2), \dots, x(t_N)\}$, then there is a direct relation between the SF and the ACF of the process. The SF is given in general by

$$\begin{aligned} SF(\tau) &= \frac{1}{N} \sum_{i=1}^N [x(t_i) - x(t_i + \tau)]^2 \\ &= \langle [x(t) - x(t + \tau)]^2 \rangle \end{aligned} \quad (\text{A1})$$

The autocovariance function of the time process $x(t)$ is given by

$$V_{x,x}(\tau) = \langle [x(t) - \overline{x(t)}] [x(t + \tau) - \overline{x(t)}] \rangle \quad (\text{A2})$$

The variance S^2 of the process $x(t)$ is equal to

$$S^2 = \frac{1}{N} \sum_{i=1}^N [x(t_i) - \overline{x(t)}]^2 = \langle [x(t) - \overline{x(t)}]^2 \rangle \quad (\text{A3})$$

note that in the denominator we have N instead of $N - 1$ (i.e. the biased estimator) due to the fact that theoretically $\overline{x(t)}$ is estimated directly from the parent distribution and not from the data set itself.

The ACF is given by

$$ACF(\tau) = \frac{V_{x,x}(\tau)}{S^2} \quad (\text{A4})$$

By expanding the terms of equation (A2) and taking into account the fact that we are dealing with a stationary process i.e. $\overline{x(t)} = \overline{x(t + \tau)} = \langle \overline{x(t)} \rangle = \text{const.}$ we get

$$V_{x,x}(\tau) = \langle x(t)x(t + \tau) \rangle - \overline{x(t)}^2 \quad (\text{A5})$$

Similarly for the variance

$$S^2 = \overline{x(t)^2} - \overline{x(t)}^2 = V_{x,x}(0) \quad (\text{A6})$$

From equation (A1)

$$SF(\tau) = \overline{x(t)^2} - 2 \langle x(t)x(t + \tau) \rangle + \overline{x(t + \tau)}^2 \quad (\text{A7})$$

By means of equation (A6)

$$SF(\tau) = 2 \left[S^2 + \overline{x(t)^2} - \langle x(t)x(t + \tau) \rangle \right] \quad (\text{A8})$$

Finally from equation (A4) and equation (A5) the SF for a stationary process is given by

$$\begin{aligned} SF(\tau) &= 2 [S^2 - V_{x,x}(\tau)] \\ &= 2S^2 \left[1 - \frac{V_{x,x}(\tau)}{S^2} \right] \\ &= 2S^2 [1 - ACF(\tau)] \end{aligned} \quad (\text{A9})$$

After equation (A9) the normalized SF (NSF) is defined as:

$$NSF(\tau) = \frac{SF(\tau)}{S^2}$$

$$= 2[1 - ACF(\tau)] \quad (\text{A10})$$

As $t \rightarrow \infty$, $ACF(\tau) \rightarrow 0$ (i.e. there is no linear correlation between the various measurements), from equation (A9) $SF(\tau) \rightarrow 2S^2$ and from equation (A10) $NSF(\tau) \rightarrow 2$.

APPENDIX B: FUNCTIONAL RELATION BETWEEN THE STRUCTURE FUNCTION AND THE POWER SPECTRAL DENSITY

Consider a zero mean stationary real time process $x(t)$ with autocovariance function (equation (A5))

$$V_{x,x}(\tau) = \langle x(t)x(t+\tau) \rangle = \langle x(t)x^*(t+\tau) \rangle \quad (\text{B1})$$

where the asterisk denotes complex conjugation. By rewriting the terms inside the mean as a function of their Fourier transform, $Y(f)$, for $-\infty < f < \infty$, equation (B1) reads

$$\begin{aligned} V_{x,x}(\tau) &= \left\langle \int_{-\infty}^{+\infty} Y(f)e^{-2\pi ift} df \int_{-\infty}^{+\infty} Y^*(f)e^{2\pi if(t+\tau)} df \right\rangle \\ &= \left\langle \int_{-\infty}^{+\infty} Y(f)Y^*(f)e^{2\pi if\tau} df \right\rangle \\ &= \left\langle \int_{-\infty}^{+\infty} |Y(f)|^2 e^{2\pi if\tau} df \right\rangle \end{aligned} \quad (\text{B2})$$

and due to stationarity, none of the terms is varying as a function of t (that is, independency of time translations e.g. Bendat & Piersol (1986))

$$V_{x,x}(\tau) = \int_{-\infty}^{+\infty} |Y(f)|^2 e^{2\pi if\tau} df \quad (\text{B3})$$

According to the first equality of equation (A9) and equation (A6) for $-\infty < f < \infty$

$$\begin{aligned} SF(\tau) &= 2[V_{x,x}(0) - V_{x,x}(\tau)] \\ &= 2 \left[\int_{-\infty}^{+\infty} |Y(f)|^2 df - \int_{-\infty}^{+\infty} |Y(f)|^2 e^{2\pi if\tau} df \right] \\ &= 2 \int_{-\infty}^{+\infty} (1 - e^{2\pi if\tau}) |Y(f)|^2 df \end{aligned} \quad (\text{B4})$$

Based on the Euler's formula, $e^{i\kappa} = \cos(\kappa) + i \sin(\kappa)$, and ignoring the imaginary part dealing only with the phases

$$SF(\tau) = 2 \int_{-\infty}^{+\infty} [1 - \cos(2\pi f\tau)] |Y(f)|^2 df \quad (\text{B5})$$

In order to estimate the variability power contained in the frequency interval between f and $f + df$, we can omit the distinction between positive and negative frequencies and regard f as varying between 0 and ∞ (Press et al. 1992). By considering the sum of the modulus-squared of the sinusoidal amplitudes $Y(f)$ and $Y(-f)$, we estimate the *one-sided* PSD of $x(t)$ (Press et al. 1992) as

$$\mathcal{P}(f) = |Y(f)|^2 + |Y(-f)|^2 \quad \text{for } 0 \leq f < \infty \quad (\text{B6})$$

Since $x(t) \in \mathbb{R}$,

$$|Y(f)|^2 = |Y(-f)|^2 \quad (\text{symmetry about the y-axis}) \quad (\text{B7})$$

and thus,

$$\mathcal{P}(f) = 2|Y(f)|^2 \quad \text{for } 0 \leq f < \infty \quad (\text{B8})$$

From the symmetric property (equation (B7)) and due to the fact that $\cos(f) = \cos(-f)$ for $0 \leq f < \infty$, equation (B5) reads

$$SF(\tau) = 4 \int_0^{\infty} [1 - \cos(2\pi f\tau)] |Y(f)|^2 df \quad (\text{B9})$$

yielding from equation (B8)

$$SF(\tau) = 2 \int_0^{\infty} [1 - \cos(2\pi f\tau)] \mathcal{P}(f) df \quad \text{for } 0 \leq f < \infty \quad (\text{B10})$$

For a PSD of a power-law form $\mathcal{P}(f) = \kappa f^{-\lambda}$, with $1 < \lambda < 3$ and κ a positive constant the last integration (equation (B10)) has an analytical solution

$$SF(\tau) = -2^\lambda \kappa \pi^{\lambda-1} \Gamma(1-\lambda) \sin\left(\frac{\lambda\pi}{2}\right) \tau^{\lambda-1} \quad (\text{B11})$$

where $\Gamma(x)$ is the (complete) Gamma function. That means that only when the following requirements

- (i) stationarity (also called *weakly stationarity* i.e. mean value and autocovariance function independent of time translations, see e.g. Bendat & Piersol 1986).
- (ii) zero mean data set.
- (iii) The frequency range f should vary from 0 to ∞ .
- (iv) The PSD should be given from a power-law form with index $1 < \lambda < 3$.

are fulfilled then there is a direct relation between SF and PSD connecting the slopes of the two quantities, β and λ respectively, $\beta = \lambda - 1$.

APPENDIX C: THE GENERAL POWER SPECTRAL DENSITY OF A SINGLE TRIANGULAR SHOT

The general PSD function of a triangular shot of the form of equation (2) is given by

$$\begin{aligned} \mathcal{P}_{\mathcal{I}}(f) &= \frac{I_0^2}{8\pi^4 t_d^2 t_r^2 f^4} \{t_d^2 + t_d t_r + t_r^2 - \\ &\quad (t_d + t_r) [t_r \cos(2\pi t_d f) + t_d \cos(2\pi t_r f)] + \\ &\quad t_d t_r \cos(2\pi(t_r + t_d)f)\} \quad \text{for } t_r \neq 0 \text{ \& } t_d \neq 0 \quad (\text{C1}) \end{aligned}$$

Figure C1 shows the form of this PSD function for the case of $t_r = 0.5$ t.u., $t_d = 100$ t.u. and $I_0 = 0.04$ flux units. It is characterized by two distinct breaks at $f_1 = t_d^{-1}$ and $f_2 = t_r^{-1}$ and the slope changes from 0 to -2 as we pass from $f < f_1$ to $f_1 < f < f_2$ and becomes even steeper for $f > f_2$ with a slope of -4. For the latter region beat frequencies appear which are separated by $\Delta f = t_r^{-1}$.

For instantaneous rise $t_r = 0$ the PSD is equal to

$$\begin{aligned} \mathcal{P}_{\mathcal{I}}(f) &= \frac{I_0^2}{8\pi^4 t_d^2 f^4} [1 + 2\pi^2 t_d^2 f^2 - \cos(2\pi t_d f) - \\ &\quad 2\pi t_d f \sin(2\pi t_d f)] \end{aligned} \quad (\text{C2})$$

for instantaneous decay the PSD is given again from the last relation (equation (C2)) having the t_d replaced with t_r .

Finally, for the case of symmetric shots $t_r = t_d = \tau$ and equation (C1) reads

$$\mathcal{P}_{\mathcal{I}}(f) = \frac{I_0^2}{\pi^4 \tau^2 f^4} \sin^4(\pi \tau f) \quad (\text{C3})$$

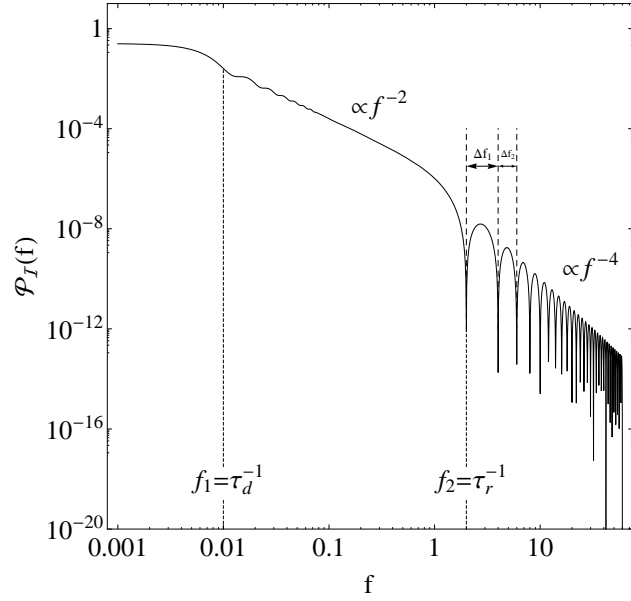


Figure C1. The general PSD function of a triangular-shaped shot (equation (2)) (in logarithmic scale). The shot parameters are $\tau_r = 0.5$ t.u., $\tau_d = 100$ t.u. and $I_0 = 0.04$ flux units. The shape of the PSD is characterized by two breaks at $f_1 = \tau_d^{-1}$ and $f_2 = \tau_r^{-1}$. For $f < f_1$ it is flat and for $f_1 < f < f_2$ it becomes steep with a slope of -2. For $f > f_2$ the PSD becomes even steeper with a slope of -4 and beat frequencies separated by $\Delta f_1 = \Delta f_2 = \dots = \Delta f_n = \tau_r^{-1}$ start to appear in the general trend.