Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **408,** 2413–2425 (2010) doi:10.1111/j.1365-2966.2010.17295.x

Fast spectral variability from Cygnus X-1

Y. X. Wu,[1,2⋆] T. M. Belloni[2⋆] and L. Stella[3]

[1]*Department of Engineering Physics and Center for Astrophysics, Tsinghua University, Beijing 100084, China*
[2]*INAF-Osservatorio Astronomico di Brera, Via Bianchi 46, I-23807 Merate, Italy*
[3]*INAF-Osservatorio Astronomico di Roma, via Frascati 33, I-00040 Monteporzio Catone, Italy*

Accepted 2010 July 2. Received 2010 July 2; in original form 2010 March 21

ABSTRACT

We have developed an algorithm that, starting with the observed properties of the X-ray spectrum and fast variability of an X-ray binary, allows the production of synthetic data reproducing observables such as power density spectra and time lags as well as their energy dependence. This further allows us to reconstruct the variability of the parameters of the energy spectrum and to reduce substantially the effects of Poisson noise, helping in studying fast spectral variations. We have applied the algorithm to the *Rossi X-ray Timing Explorer* data of the black hole binary Cygnus X-1, fitting the energy spectrum with a simplified power-law model. We found that the distribution of the power-law spectral indices on time-scales as low as 62 ms is limited between 1.6 and 1.8. The spectra index is positively correlated with the flux even on such time-scales.

Key words: methods: data analysis – methods: statistical – X-rays: binaries – X-rays: individual: Cygnus X-1.

1 INTRODUCTION

Black hole binaries (BHBs) exhibit considerable X-ray variability on a wide range of time-scales. The study of X-ray fast time variability has become an important astrophysical research tool that helps us gain better insight into the physical process at work near the black hole (see the recent review of van der Klis 2006). For instance, the dynamic time-scale for the motion within a few Schwarzschild radii of a 10-$M_\odot$ black hole is of the order of milliseconds. Further considering that most of the gravitational energy of accretion matter is released in the inner area of a few Schwarzschild radii, the variability at short time-scales can be used to probe the accretion-flow dynamics and geometries within the strong-field region.

Time variability can be studied in the time domain or in the frequency domain. The latter is based on the Fourier transform (FT) and is usually based upon two basic techniques: the power density spectrum (PDS) and the time-lag spectrum. The square of FT amplitudes as a function of Fourier frequency constitutes the PDS, which provides the estimate of variance at different frequencies. The time-lag spectrum is obtained from the phase lag, i.e. the phase angle difference between the Fourier vectors at different energy channels. In practice, the PDS and the lag spectra are usually averaged over many segments of observations and frequencies in order to increase the statistical significance. The FT is reversible; the time series can be reconstructed from its FT by means of the inverse Fourier transform (IFT). In contrast, the PDS is not reversible, since the phase information in the FT is lost. In principle, there is an infinite variety of different signals that will yield the same PDS.

The fast variability observed from BHBs is of stochastic nature and as such cannot be modelled directly. In other words, it is not possible to reproduce the exact observed variations. The aim of time-series analysis is to characterize the average properties that give rise to the fluctuations, under the assumption that the process is stationary. A successful model should reproduce the PDS and the lag spectrum, as well as other statistical properties of the signal (see e.g. Uttley, McHardy & Vaughan 2005). A conventional model describing the temporal fluctuation is the shot-noise model (Terrell 1972; Negoro, Miyamoto & Kitamoto 1994). It has become clear, however, that in this framework complex shot profiles or distributions of shot durations and amplitudes have to be assumed to model the variability of BHBs (e.g. Miyamoto et al. 1988; Belloni & Hasinger 1990; Lochner, Swank & Szymkowiak 1991). An alternative way is to apply linear state space models which are based on stochastic processes or on autoregressive processes to describe the temporal variability (König & Timmer 1997; Pottschmidt et al. 1998). Uttley et al. (2005) use a non-linear model to explain the log-normal flux distribution and rms–flux relation. All these models are phenomenological; based on the PDS alone they try to reproduce the observed properties through a mathematical model, which can provide constraints on physical models. On the other hand, Arévalo & Uttley (2006) attempted a more physically constrained generating process to model all the spectral-timing properties simultaneously.

The usual course of action is to extract information in the frequency domain, such as a PDS, from the time series. However, in

⋆E-mail: wuyx@mails.thu.edu.cn (YXW); tomaso.belloni@brera.inaf.it (TMB)

some cases we need to do the opposite, e.g. to reconstruct the time series from the PDS. The simulation of random time series with an arbitrary PDS has a long and established history (see e.g. Davis, Hagan & Borgman 1981; Liu & Munson 1982) in the field of digital signal processing. There are also papers on simulating time series with specific marginal distribution, e.g. lognormal (Johnson 1994). In astrophysical research, the reasons and benefits to perform such a reconstruction are various. Sometimes it provides a more direct tool to judge the models or simulation methods (for an example, see Timmer & König 1995). Also it can be used to estimate the error bars by the Monte Carlo simulation (e.g. Done et al. 1992). Since the PDS does not contain phase information, so in order to obtain the time series one must assign values to the phases as the reconstruction based on the PDS alone is not unique. One easy way to generate data that reproduce a given PDS is to choose the Fourier amplitude according to the PDS and assign random phases between $[0, 2\pi]$ (Done et al. 1992). Based on the theory of linear stochastic process and the fact that the PDS itself follows a χ^2 distribution, Timmer & König (1995) proposed an algorithm practically identical to that of Davies & Harte (1987) in order to produce a variety of possible non-deterministic linear time series from the PDS by randomizing both the phases and amplitudes. Some authors use the energy-resolved PDS instead of the total PDS given by the method of Timmer & König (1995) and shift the phase to yield light curves with a desired lag between energy bands (e.g. Zoghbi et al. 2010). Recently, the non-linear behaviour of observed light curves was studied by Uttley et al. (2005). They suggested that an additional exponential transform needs to be applied to the time series created with the method of Timmer & König (1995).

BHBs are known to exhibit X-ray spectral evolution on short time-scales. This evolution is reflected in the presence of lags between the light curves at different energy ranges and asymmetries of the cross-correlation function between them, as well as fast variations of the corresponding hardness ratio. However, the conventional spectral models applied to these systems are designed to fit the energy spectrum averaged over, usually, several thousand seconds. Because of limited statistics, it is not possible to follow the energy spectrum over the short time-scales corresponding to the observed fast variability.

In this work, we propose a new technique that simulates a time series starting from the actual intensity measurements. Specifically, we simulate the light curves in different energy bins reproducing all the properties observed in the real data: the average energy spectrum, the PDS as a function of energy and the frequency-dependent lag spectra between different energy bands. In the simulation, no Poisson noise is introduced. With the simulated 'clean' light curves in different energy bins, we can study the variations of the energy spectrum on short time-scales. Our work is improved (or different) in three aspects when compared with the papers mentioned above. First, besides the PDS, we make use of other measurements, including energy spectra and lag spectra as input to the simulation. Secondly, we require that the simulated time series should reproduce almost all the timing and spectral properties. Thirdly, we explore the possible application of the method, including filtering Poisson noise and data extrapolation. Our work, based on those by Timmer & König (1995) and Uttley et al. (2005), can be seen as a continuation of them. In contrast to Arévalo & Uttley (2006), our work is model-independent and aims at developing an algorithm that recovers the time series preserving the information and filtering the noise.

This paper is organized as follows. The initial works of data analysis and parameter estimation are introduced in Section 2, including the non-linearity study (Section 2.1), energy dependence of PDS (Section 2.2), lag spectra (Section 2.3), coherence function (Section 2.4) and distribution of the phases (Section 2.5). The important results are presented in Section 3. In Section 3.1, the algorithm is defined step by step. The simulated light curve is compared with the observed one in Section 3.2 and the issue of Poisson noise subtraction is discussed in Section 3.3. As one important application, the energy spectra on short time-scales (dynamic energy spectra) are derived and the variation of the spectral shape is investigated in Section 3.4. Another possible application, the simulation of data with better time and energy resolution than observations, is discussed in Section 3.5. The issues associated with phase and noise are discussed in Section 3.6. Conclusions are presented in Section 4.

2 DATA ANALYSIS AND PARAMETER ESTIMATION

The central idea of the algorithm is to synthesize light curves that reproduce all observed properties in both the time and the frequency domains. These are

(i) the average count rate in different energy bands, both integrated (spectrum) and as a function of time (light curves);
(ii) the relation between rms variability and count rate in the light curves at different energies (rms–flux relation);
(iii) the shape and normalization of the PDS as a function of energy;
(iv) the phase/time-lag spectrum as a function of energy.

Once these properties are extracted from the data, a synthetic data curve can be constructed to reproduce the same results.

A preliminary step is to obtain the above quantities with good accuracy for a BHB. We used a single observation of Cygnus X-1 from the proportional counter array (PCA) on board the *Rossi X-ray Timing Explorer* (*RXTE*; Obs ID 10238-01-05-000). Cyg X-1 is the first discovered BHB and has been studied for several decades. Its brightness and persistence make it a perfect target for X-ray timing research. The observation was carried out in 1996 March, when Cyg X-1 was in its hard spectral state, the most common for the source (e.g. Wilms et al. 2006). The data configuration used here is the generic PCA binned mode B_16ms_64M_0_249, which provides high resolution in both energy (64 channels over the full 2–60 keV PCA band) and time (16 ms). This is the main reason why we selected this particular observation, which also has the advantage of being made with all five proportional counter units of the PCA, increasing the source count rate. In order to obtain sufficient statistics for the analysis, we rebinned the data into eight energy bins between 0.14 and 25 keV. In this way in each energy bin the mean count rate is above $\sim$1000 counts s^{-1}. Note that the contribution of background photons to each of these eight bins is minor compared to the source counts.

For the analysis, we used custom-made software written in the IDL environment. For each of the eight energy bins, we extracted a Fourier spectrum from data stretches of a length of 128 s, up to a Nyquist frequency of 32 Hz. These Fourier spectra were averaged for constructing the PDS (normalized to the squared fractional rms; see Belloni & Hasinger 1990) and the phase-lag spectra. The Poisson contribution was subtracted by using *RXTE* recipes (see Zhang et al. 1995).

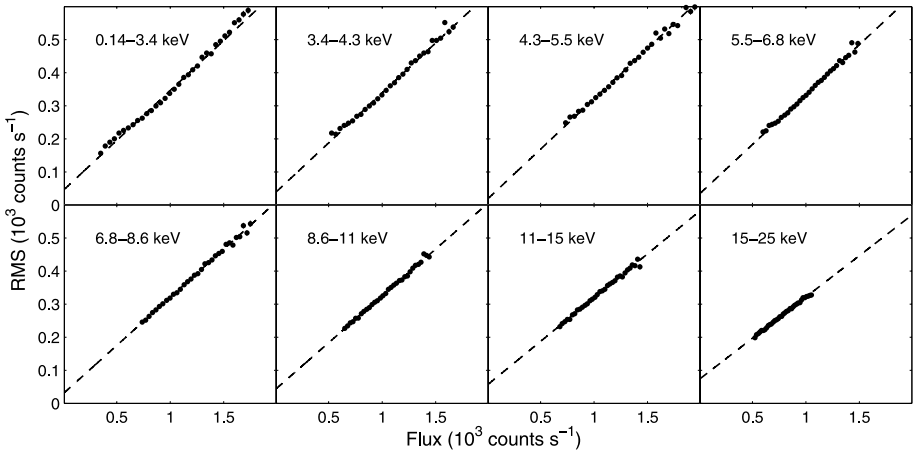

Figure 1. The rms–flux relations for eight energy bins. They are produced by binning the 1–32 Hz rms measured in 1-s segments into flux bins. The dashed line is the best-fitting linear model for each plot.

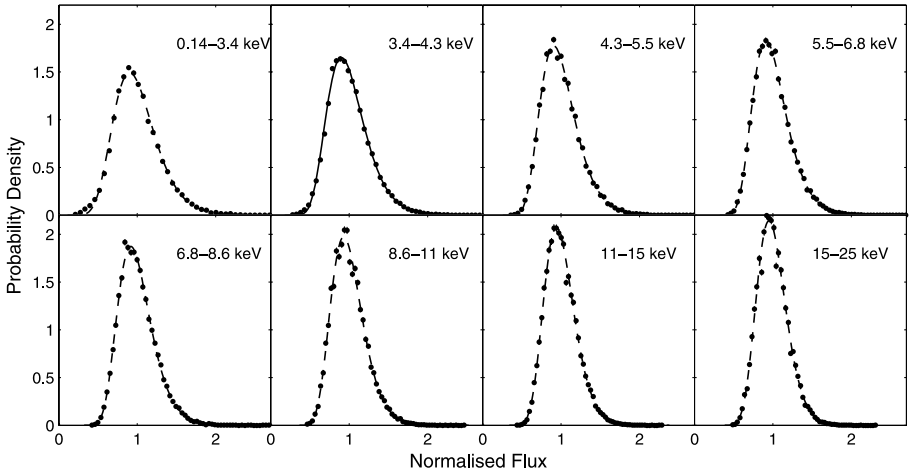

Figure 2. The flux distribution expressed as a probability density, i.e. the data points per flux bin normalized by the flux bin width and by the total number of data points. The flux is calculated over 0.25-s time bins. The dashed line is the best-fitting lognormal model for each plot.

2.1 Rms–flux relation

Uttley et al. (2005) showed that the rms–flux relation, the non-linear behaviour and the lognormal flux distribution observed in the hard state of BHB represent three different aspects of the same underlying process. This non-linearity can be reproduced as an exponential of a linear light curve. As the plan is to simulate the source light curve in different energy bands, we must first check that this non-linearity also holds for a separate energy band. We then produced the rms–flux relation and the flux (count rate) distribution for each of our eight energy bins (see Figs 1 and 2).

The rms in Fig. 1 was measured by integrating the PDS of the eight bins over the 1–32 Hz frequency interval for 1-s segments. Its relation to the count rate, also in 1-s bins, is consistent with linearity for all bins. This linear relation also holds when the length of the light-curve segments and the frequency range for the integration are changed. At the same time, the flux distribution (with a time bin of 0.25 s in Fig. 2) fits a lognormal model. Further subdividing the energy range into narrower bins, we did not find significant deviations.

As shown by Uttley et al. (2005), for typical observed light curves with the PDS dominated by broad components and a fractional rms of 20–40 per cent, the distorting effect of the exponential transformation on the shape of the PDS is relatively small. A quantitative analysis (not presented here) suggests that for Lorentzian-shaped PDS components with fractional rms smaller than 50 per cent, the distortion is not serious if the quality factor[1] $Q \lesssim 2$. We tested the distortion on the PDS caused by the exponential transformation of data in the time domain: we calculated the PDS from the logarithm of the real data and compared it with the original one (Fig. 3). The PDS shapes are almost unchanged between the raw data and the logarithmically transformed data. Therefore in our simulation below, it is justified to use the observed PDS as the PDS of the input linear light curve without the need to correct the distortion effect of the exponential transformation. However, it is still important to apply the appropriate correction to the normalization of the input PDS in order to reach the desired variance in the output light curve, because the exponential transformation will cause an increase in the light-curve variance. Also the mean count rate of the light curve would change after the exponential transformation, and a correction factor needs to be multiplied by the non-linear light curve.

[1]The quality factor is the ratio of the Lorentzian centroid frequency to the FWHM.

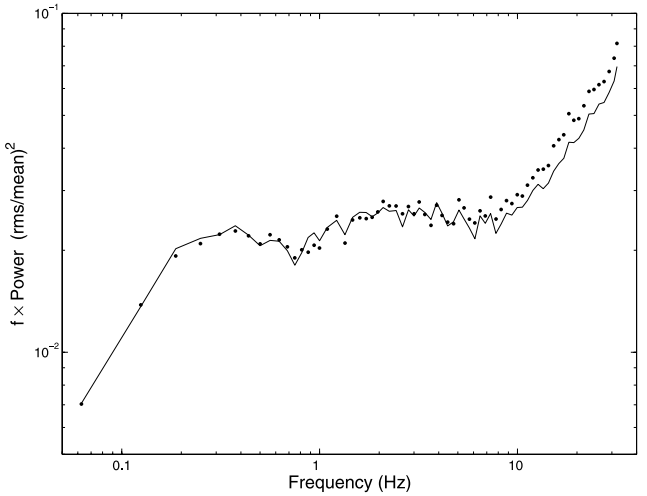

Figure 3. The PDS of the raw data (solid line) and the logarithmically transformed data (dot). The latter is renormalized to be directly compared with the former. The Poisson noise is not subtracted from the PDS.

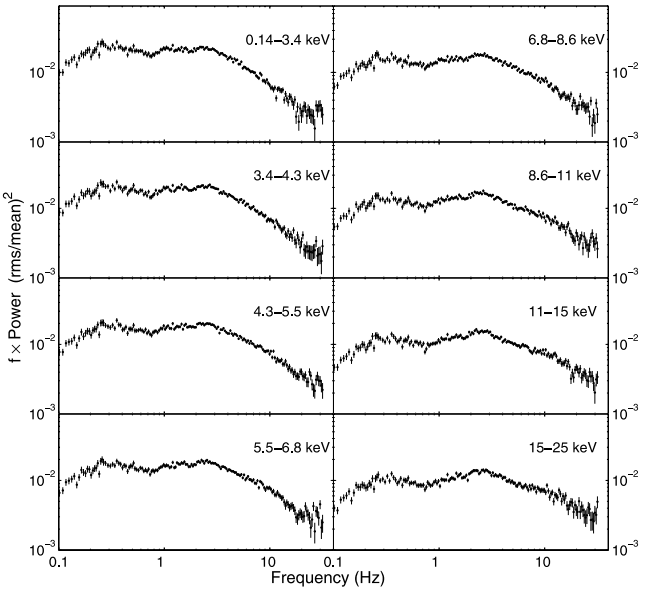

Figure 4. The PDS in eight energy bins, with the Poisson noise subtracted.

2.2 Energy dependence of the PDS

We extracted an average PDS from each of the eight energy bins, covering the frequency range of 0.008–32 Hz (Fig. 4). No narrow quasi-periodic oscillations are seen. A simple model consisting of two broad Lorentzians was used for the fit. The goodness of the fit is reasonably good, with all reduced χ^2 smaller than 2.2 (obtaining a formal reduced χ^2 of the order of unity is difficult for these high-signal PDS). Adopting a more complex model (e.g. Belloni et al. 1997), the goodness of fit would be improved but the fit parameters would be poorly constrained. There are three free parameters for each Lorentzian: the normalization (the square of the integrated fractional rms), the centroid frequency and the full width at half-maximum (FWHM). The evolution of the PDS shape with energy can be well described by the energy dependence of the best-fitting parameters, which is shown in Fig. 5. In the case of the second Lorentzian, for the last three energy bins the best-fitting centroid

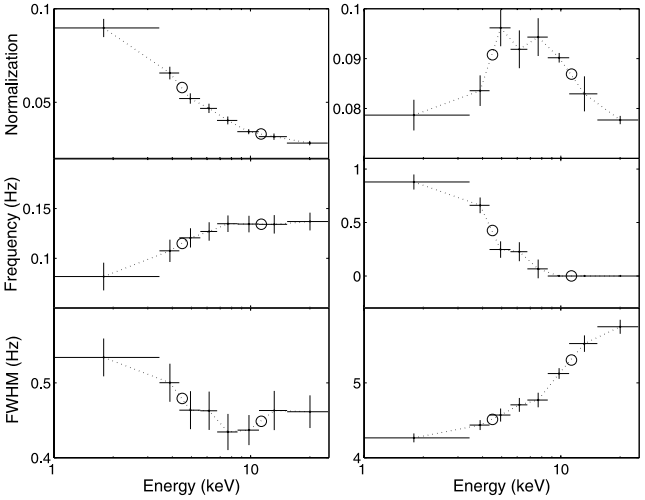

Figure 5. The best-fitting parameters of the two-Lorentzian model for the PDS at different energies. The left-hand and right-hand columns correspond to the two Lorentzian components. The parameters are (from top to bottom) normalization, centroid frequency and FWHM. The circles are the interpolation values between energy bins (see Section 3.5).

frequency decreases to zero, the lower bound of this free parameter. For these three bins, uncertainties were not plotted.

2.3 Energy dependence of the time-lag spectrum

The time lag of the light curve in each energy bin relative to that of the lowest energy bin (0.14–3.4 keV) was calculated from the cross-spectra between 0.06 and 30 Hz. Positive lags here correspond to the hard time series lagging the soft. The lags were logarithmically rebinned in frequency in order to reduce noise. Since their calculations involve the splitting of the data into two energy bands, compared with the PDS the measurement of time lags is more sensitive to counting noise. Nowak et al. (1999) estimated the expected noise level for the time-lag measurements and concluded that for frequencies below ∼0.1 and above ∼30 Hz, lags cannot be measured because of noise limitations. As the frequency approaches ∼0.1 or ∼30 Hz, the lags tend to zero due to the effect of noise. When sampling fluctuations become comparable to the intrinsic lags, they scatter around zero and exhibit negative values.

We adopted the same strategy as for the PDS to quantitatively describe the energy dependence of the lag spectrum in a uniform way. Nowak et al. (1999) showed that the time lags approximately show a power-law dependence upon frequency ($\propto f^{-0.7}$). We found significant deviations from a simple power-law model. The time-lag spectra show a two-humped shape similar to that published in previous studies (e.g. Miyamoto et al. 1992; Cui et al. 1997; Nowak et al. 1999). We used a two-Lorentzian model to fit the time-lag spectra. The time-lag spectra with the best-fitting two-Lorentzian models are shown in Fig. 6. Because the negative lags have small values and appear in the frequency range where the noise level dominates, they are not expected to be intrinsic. The negative lags were therefore excluded from the fit. The evolutions of the best-fitting parameters with energy are shown in Fig. 7.

2.4 Coherence function

The coherence function is a Fourier-frequency-dependent measure of the linear correlation between time series measured

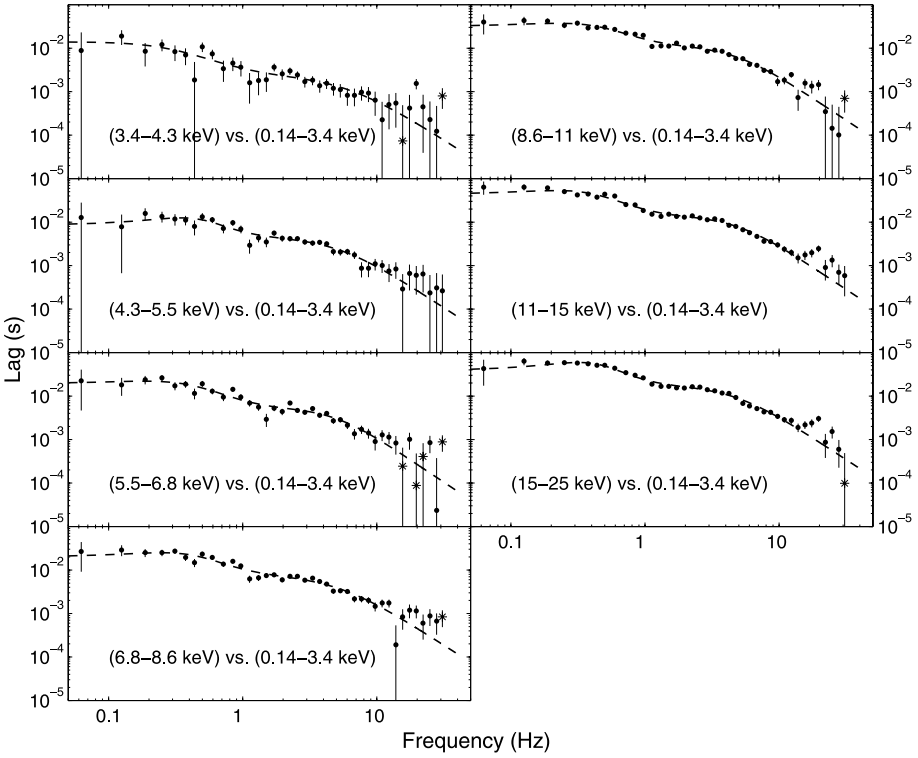

Figure 6. Time-lag spectra for various energy bins versus the lowest energy bin (0.14–3.4 keV). Dots represent the positive lags (hard lagging the soft) and asterisks represent negative lags (soft lagging the hard). The dashed lines are the best-fitting two-Lorentzian models for the positive lags.

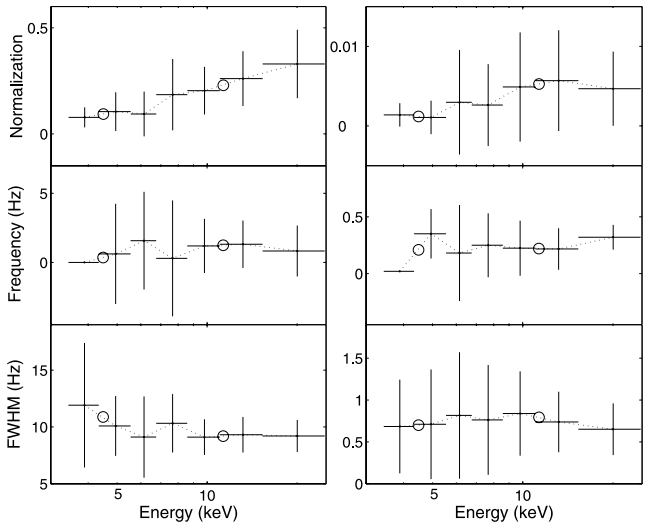

Figure 7. The best-fitting parameters of the two-Lorentzian model for time-lag spectra at different energies. The left-hand and right-hand columns correspond to the two Lorentzian components. The parameters are (from top to bottom) normalization, centroid frequency and FWHM. The circles are the interpolation values between energy bins (see Section 3.5).

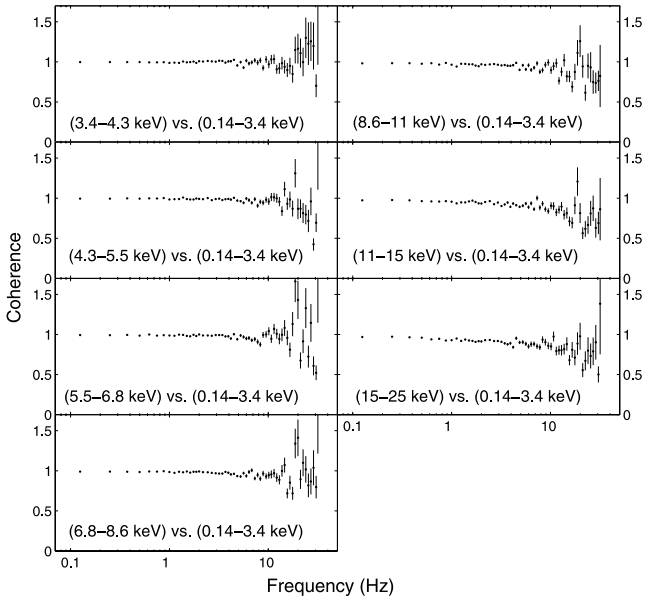

Figure 8. The coherence function for various energy bins versus the lowest energy bin (0.14–3.4 keV).

simultaneously in two energy bands (Vaughan & Nowak 1997). Our simulation does not include any incoherent variability, i.e. variations in one energy band that are not correlated with variations in other bands. In other words, the algorithm contains an underlying assumption of a single emission component in different energy bands with a single delay at a given frequency and unity coherence. We calculated the coherence function of Cyg X-1 data with correction for counting noise, following the recipe presented in Vaughan & Nowak (1997). The results are shown in Fig. 8, which demonstrate a remarkably high coherence (close to unity) over a wide frequency range and consistent with the previous coherence study of Cyg X-1 (e.g. Cui et al. 1997; Vaughan & Nowak 1997; Nowak et al. 1999). Therefore, we can say that below ∼10 Hz, the flux in each energy band can be regarded as originating from one single coherent component, whose intrinsic phase delay is indicated by the lag spectrum. The coherence becomes slightly lower at higher

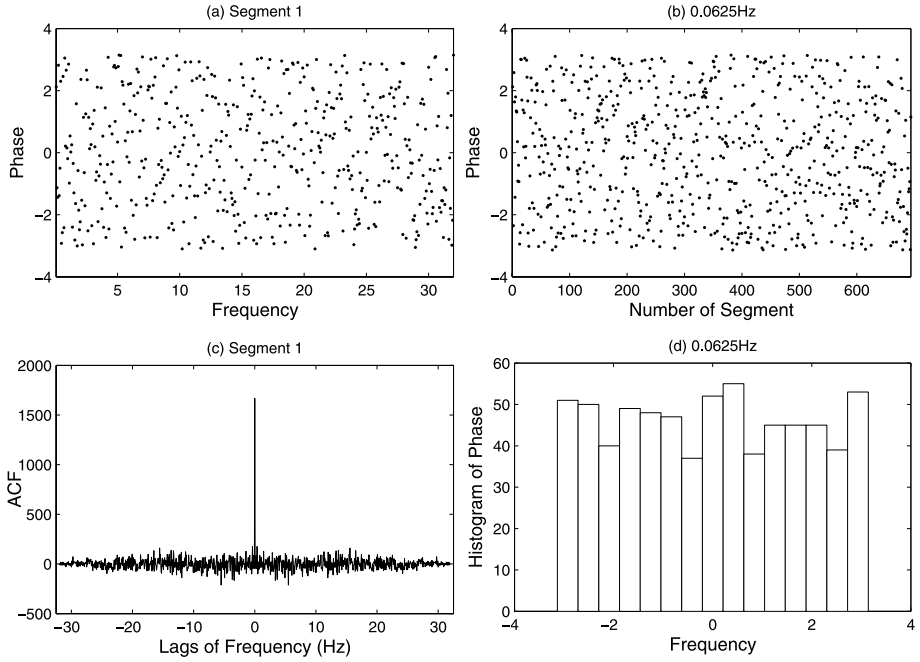

Figure 9. (a) The phase at different frequencies for the first segment, (b) the phase at 0.0625 Hz for 694 segments, (c) the autocorrelation function of phase at different frequencies for the first segment and (d) the histogram of phase at 0.0625 Hz.

frequencies, which may indicate the presence of incoherent components. We do not attempt to add them to the simulation because it is a laborious and model-dependent process and beyond the scope of this work.

2.5 Phase

In order to reconstruct the time series from the PDS and time-lag spectra, a prior phase distribution has to be assumed, since the PDS does not contain phase information. We first analysed the real data in order to derive a reasonable phase distribution. We split the 0.015 625-s binned light curve into 694 segments, each with a length of 1024 points (16 s). For each segment, we produced an FT, leading to 694 values of the phase angle for each frequency between 0.0625 and 32 Hz. Obviously, the phases between separate segments are comparable only after considering the additional phase shift caused by the time delay between their start time. If $\varphi_{j,i}$ is the phase angle at frequency f_j of the ith segment, $t_{0,i}$ and $t_{0,1}$ are the start times of the ith segment and the first segment, respectively; the 'absolute' phase at this frequency for the ith segment can be calculated as

$$\varphi'_{j,i} = \varphi_{j,i} + 2\pi f_j(t_{0,i} - t_{0,1}).$$

In Fig. 9, we plot the phase at different frequencies for the first segment (panel a), the phase at a certain frequency for different segments (panel b) and its histogram of occurrence (panel d). Moreover, we studied the autocorrelation function of the phase at different frequencies (shown in panel c). All these results clearly show that the phase follows a uniform distribution between $-\pi$ and π and the phases at different frequencies are random and independent. Therefore, in our synthesis algorithm we generate uniformly distributed random numbers in the interval $(-\pi, \pi]$ as phase angles for the FT.

3 RESULTS

3.1 The algorithm

Having obtained the energy-resolved PDS (with Poisson noise subtracted) and the time-lag spectra for our eight energy bins, and the average energy spectrum, we followed the procedure given below to generate a synthetic light curve.

(i) *Step 1*. For the lowest energy bin (bin 1), we generated uniformly distributed random numbers between $(-\pi, \pi]$ to be used as phase angle $\varphi(f_i)$ at Fourier frequency f_i ($f_i \geq 0$). In order to obtain real values for the time series, we chose the phase for the negative frequencies as $\varphi(-f_i) = -\varphi(f_i)$. For energy bins 2 through 8, the phases were reconstructed from the bin 1 values according to the measured phase-lag spectra which provide the phase shift relative to the lowest energy bin at each frequency.

(ii) *Step 2*. The amplitude of the FT at each frequency was obtained from the PDS. In order to account for the effects of the exponential transformation to the variance of the light curve, the PDS should first be renormalized in order to obtain the desired variance. For a PDS $P(f)$ in units of (rms/mean)^2 Hz^{-1} and with frequency bin size Δf, the desired fractional rms is $R^2 = \Sigma P(f)\Delta f$. The PDS must be multiplied by a factor of $\log(R^2 + 1)/R^2$ (for details, see Uttley et al. 2005). The square root of the renormalized PDS is the amplitude of the FT $A(f_i)$. The series also need to be expanded to negative frequencies with $A(-f_i) = A(f_i)$.

(iii) *Step 3*. For each energy band, we calculated the IFT of $A(f_i)\exp[j\varphi(f_i)]$ (where j is the imaginary unit) to obtain the linear time series $l(t)$ and then calculated its exponential. In order to ensure that the simulated light curve has the desired mean count rate $C(E)$ measured in the average energy spectra, a factor of $C(E)\log(R^2 + 1)/R^2$ needs to be multiplied by $\exp[l(t)]$.

(iv) *Step 4*. The time series obtained with the previous steps was stored as one light-curve segment. Steps 1–3 were then repeated to produce multiple segments.

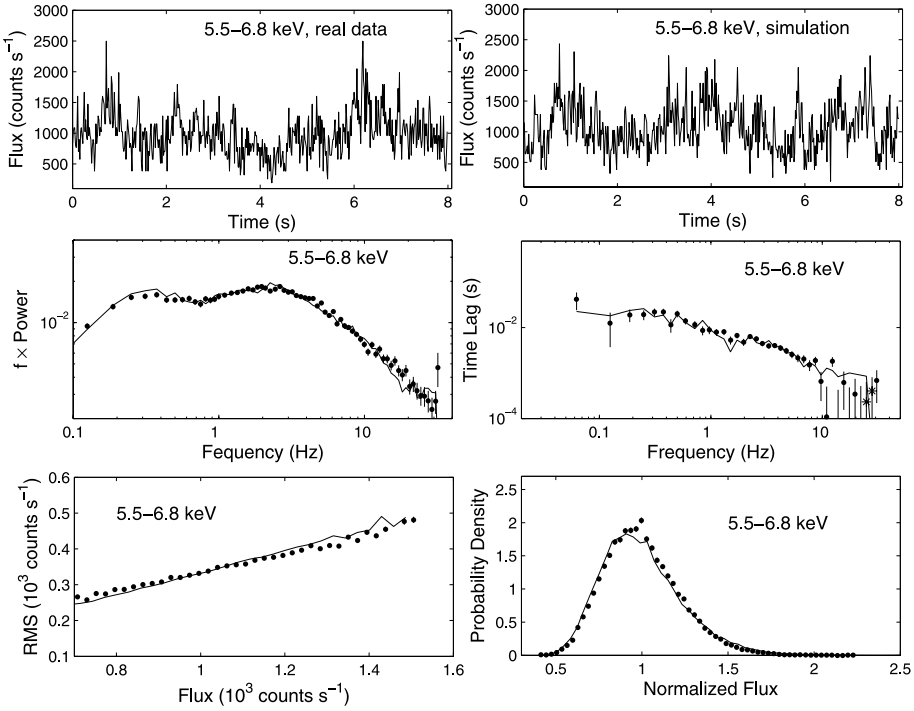

Figure 10. Comparison of the simulation with the algorithm described in Section 3.1 (with additional Poisson noise) and the real data (Obs ID 10238-01-05-000) for the energy bin of 5.5–6.8 keV. Top: one segment (1024 points) of light curve of the real data (left) and the simulation (right). Middle: the PDS (left) and the time-lag spectra (right) of the simulation (dot for positive values and asterisk for negative values) and the real data (solid line). Bottom: the rms–flux relation (left) and flux distribution (right) of the simulation (dot) and the real data (solid line). The error bars of the simulation data are plotted.

3.2 Test of the simulation

To check whether the simulated light curve replicates all the observed properties of original real data, we compared their PDS, time-lag spectra, rms–flux relation and lognormal flux distribution. The comparison relative to the energy bin of 5.5–6.8 keV is shown in Fig. 10 as an example. The simulated light curve of course cannot have exactly the same evolution as the real one, as a random-number input is involved, but it appears to be similar in the amplitude and time-scale of variance. The PDS, the time-lag spectrum, the rms–flux relation and the flux distribution are consistent with those from the real data, showing that our simulation reproduces accurately the intrinsic properties of the real data. In other words, our algorithm can synthesize data whose statistical properties are indistinguishable from those observed from Cyg X-1.

3.3 Poisson noise subtraction

The expected influence of Poisson fluctuations in the time series is represented as a white noise component in the PDS. Since it is independent of the source signal (apart from dead-time effects), the Poisson noise can be considered as a 'background' component in the PDS, against which we try to observe other features caused by the intrinsic variability of the X-ray source. If the light curve is a series of contiguous time bins, the expected Poisson noise level is simply 2 for the Leahy normalization (Leahy et al. 1983). In this work, the PDS is normalized in units of (rms/mean)^2 Hz^{-1} (Belloni & Hasinger 1990), and the expected Poisson noise level in the PDS is given by

$$P_{\rm noise} = \frac{2(C+B)}{C^2},$$

where C and B are the mean source count rate and background count rate, respectively. See Vaughan et al. (2003) for more details about the different normalizations of PDS and the corresponding Poisson noise levels. Note that the shape and level of the Poisson noise contribution to the PDS are modified by dead-time effects (for the *RXTE*/PCA, see Zhang et al. 1995).

Therefore, we can easily subtract the Poisson noise level from the PDS and obtain the 'clean' light curve without Poisson noise with our algorithm. In other words, our synthetic algorithm can be used as a filter of Poisson noise. We can check this by adding Poisson fluctuations to an initial simulated 'clean' light curve and then filter the 'dirty' light curve with our algorithm to see whether the filtered light curve resembles the initial one or not. Because the algorithm cannot repeat the exact shape of the light curve due to the phase randomization, in order to make a direct comparison between the clean light curve and filtered light curve, we have to record the phase information of the initial data as the input of the algorithm instead of using random phases. We do this here because our aim is simply to check the effect of Poisson noise filtering for the synthetic algorithm.

The results are shown in Fig. 11. We can see how the Poisson noise is removed in the filtered light curve. A quantitative measurement is the standard deviation, which is 320, 420 and 340 for the clean, dirty and filtered light curves, respectively. The excess variation of the filtered light curve at high frequencies is probably caused by the distortion introduced by the exponential transform, which tends to exaggerate the positive variation, e.g. the amplitudes of the flares. The conclusion is that the synthetic light curve created by our algorithm can be considered essentially free of Poisson noise.

It is worth to point out that if the phases were known, we could reconstruct the time series strictly by IFT, because this time-to-frequency transition is completely reversible. The exponential

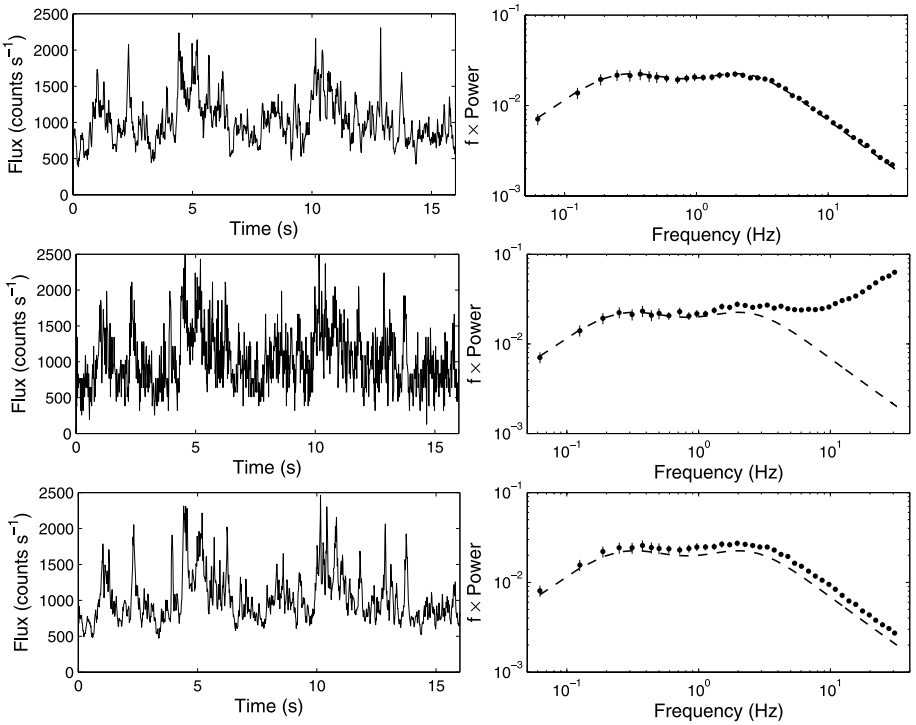

Figure 11. Top: the initial clean light curve and its PDS. Middle: the light curve with additional Poisson noise and its PDS. Bottom: the light curve filtered by our algorithm and its PDS. The dashed line in the right-hand panels is the PDS of the initial clean light curve. The PDS are calculated from 50 segments, each with a length of 1024 points. Each of the left-hand panels shows only one segment of the corresponding light curve.

transform will not be necessary. We therefore face an interesting problem: if the phase is known, the exponential is redundant for reconstructing the initial time series, which is completely defined by the Fourier spectrum; however if we know nothing about the phases and assume them to be random, in order to reproduce a time series satisfactorily we have to apply the exponential transformation. We will discuss this problem in Section 3.6.

The above process of subtracting the Poisson noise is similar to the Wiener filtering. The Wiener filter is the optimal filter in the least-squares sense for the removal of noise from a time domain signal. The Wiener filter is designed in the Fourier domain and can be expressed as (Press et al. 1992)

$$\Phi(f) = \frac{|S(f)|^2}{|S(f)|^2 + |N(f)|^2}$$

in which S and N are the FTs of the intrinsic signal and the noise, respectively. The denominator $|S(f)|^2 + |N(f)|^2$ is proportional to the PDS of the measured light curve (under the assumption that signal and noise are statistically independent). The filter can be constructed if the true form of the intrinsic power $|S(f)|^2$ is known or can be estimated well. Our algorithm and the Wiener filter share some common ideas – to separate noise and signal in the frequency domain, which cannot be done in the time domain. We fixed the noise power to 2 and applied a Wiener filter to the same data in Fig. 11. The difference between the Wiener filtering solution and ours is shown in Fig. 12. The filtering effect is generally similar for the two methods. If the PDS used in the Wiener filtering and our algorithm are those averaged over many segments (panels *b* and *d* in Fig. 12), our algorithm appears to preserve more short time-scale fluctuations, which makes it more similar to the original one. If the PDS of the single segment shown in the figure is used in designing the Wiener filter (panel *c* in Fig. 12), its solution would be much more noisy than the original light curve. From this

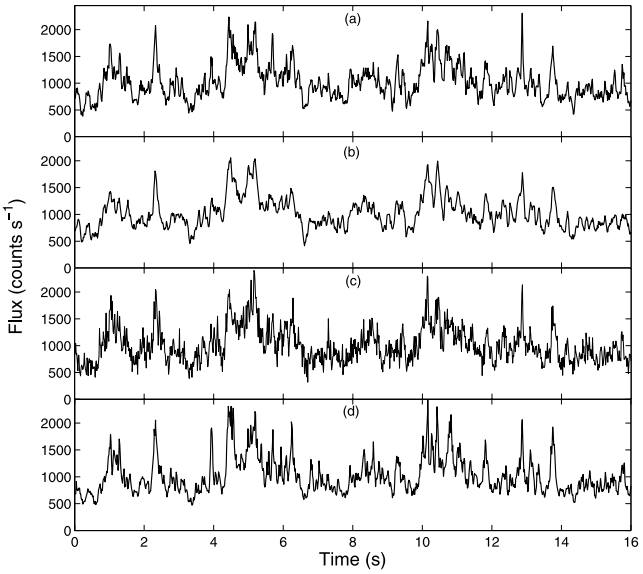

Figure 12. (a) The initial clean light curve (one segment, the same as in the top-left panel of Fig. 11). This is used as the input for both Wiener filtering and our algorithm, after adding Poissonian noise (see the text). (b) The Wiener filtering solution with the averaged PDS over many segments. (c) The Wiener filtering solution with the PDS of the single segment. (d) The solution of our algorithm (the same as in the bottom-left panel of Fig. 11).

perspective, our algorithm seems to provide a more stable filter as it is able to subtract the noise and at the same time avoid destructing too much rapid variability. This is probably due to the exponential transformation that can restore the fluctuation amplitude to a certain

extent. The price of the exponential transformation is that the exact shape of the light curve is slightly distorted.

3.4 Dynamic energy spectra

If the photon count in one bin of the energy spectrum is N, the relative standard deviation expected from the Poisson distribution is $1/\sqrt{N}$. When we want to obtain the average energy spectra with high statistical significance, we need a sufficiently long exposure (hundreds to thousands of seconds) to accumulate enough photons. It is important to see how the energy spectra evolve on short time-scales which, however, in this way is not possible. The energy spectra on short time-scales, which we name dynamic energy spectra, can be obtained by aligning the light curves of different energy bins in time and obtaining the counts in every time bin as a function of energy. The problem of this analysis from the real data is that the Poisson fluctuation is severe in this case due to the small number of counts. The simulated light curve obtained with our synthetic algorithm, as presented in Section 4, does not include Poisson noise. The dynamic energy spectrum produced from the simulated light curves is therefore 'cleaner' and can reveal the underlying properties otherwise hidden by noise.

The dynamic energy spectra with eight energy bins were calculated for both real data and simulation at three time-scales (or time bin sizes), 0.0625, 0.25 and 1 s, and then fitted with XSPEC using the PCA detector response matrix. Note that the synthetic data, having reproduced the background-subtracted energy spectrum, are also background free. The lowest energy bin (0.14–3.4 keV) is excluded due to the uncertainty in the PCA calibration below 3 keV and a simple power law is fitted to each of the dynamic energy spectra. There are in total 25 600, 6400 and 1600 dynamic energy spectra that were fitted for the three time-scales, respectively. The power-law photon index Γ is the parameter that we studied to reflect the basic shape of dynamic energy spectra. For a comparison of real data and simulation, we plot the time evolution of Γ as well as the flux covering the whole energy band (0.14–25 keV; Fig. 13), the correlation between Γ and flux (Fig. 14) and the histogram of Γ (Fig. 15). The noise-free reconstructed data provide a 'cleaner' view of the Γ distribution and the Γ–flux correlation. One possible reason for the improved correlation is that the simulation does not include any incoherent variations that may weaken the correlation for the real data. However, we have shown that the coherence in the real data is very close to unity for most of the frequencies considered here and therefore this possibility can be excluded.

Combining the above results, we can conclude as follows.

(i) The correlation between Γ and flux is somewhat higher for the simulation than for the real data.

(ii) The distribution of Γ is narrower for the simulation than for the real data.

(iii) The above differences between simulation and real data tend to increase at shorter time-scales, i.e. for lower photon count numbers.

The correlation between Γ and flux is consistent with the previous results that the hardness ratio anticorrelates with the X-ray flux (e.g. Cui, Feng & Ertmer 2002; Liu & Li 2004; Wilms et al. 2006) or that the photon index correlates with the flux on a time-scale of days (e.g. Zdziarski et al. 2002; Gierliński & Zdziarski 2003; Pottschmidt et al. 2003). The significantly better correlation for simulated data (Fig. 14) shows that our algorithm enables us to reduce the effects of

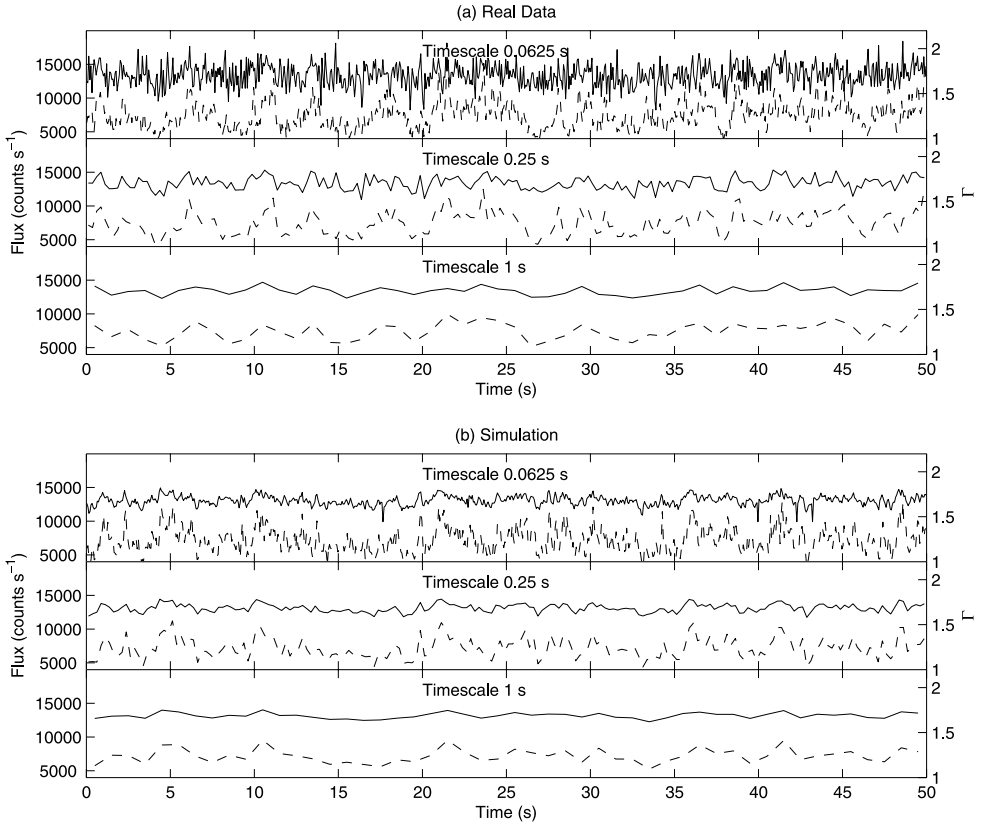

Figure 13. The 0.14–25 keV light curve (dashed line) and the corresponding power-law photon index Γ evolution (solid line) for (a) real data and (b) simulation (without Poisson noise), with time bin sizes of 0.0625, 0.25 and 1 s from top to bottom.

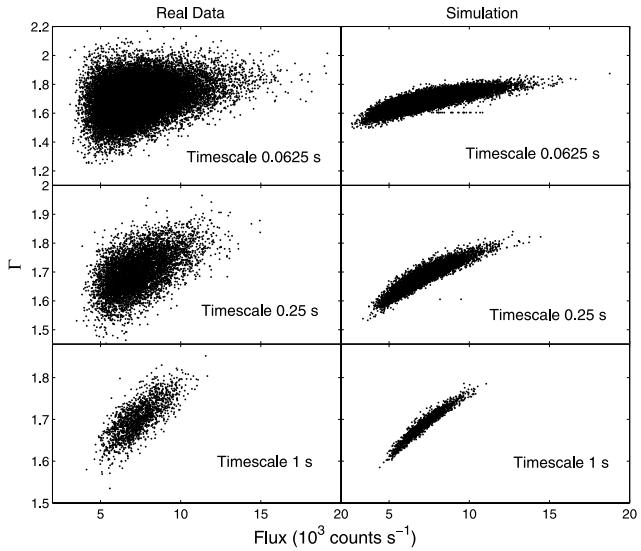

Figure 14. The correlation between the power-law photon index Γ and flux at three time-scales (or time bin sizes) of 0.0625, 0.25 and 1 s for real data (left-hand column) and simulation (right-hand column).

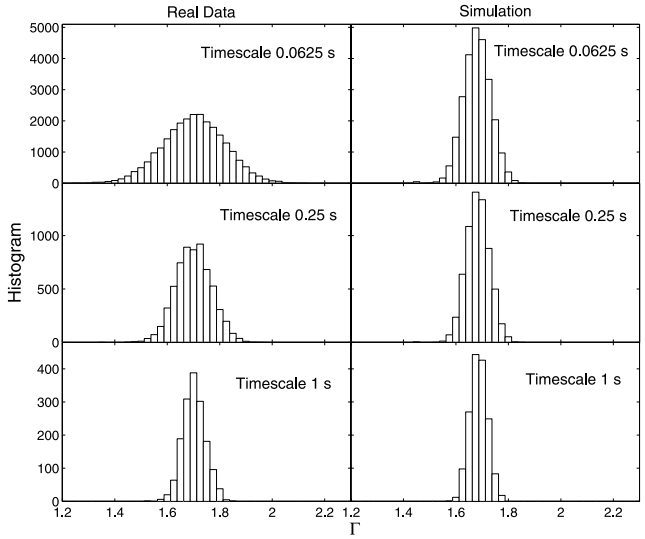

Figure 15. The histogram of the power-law photon index Γ at three time-scales (or time bin sizes) of 0.0625, 0.25 and 1 s for real data (left-hand column) and simulation (right-hand column).

Poisson noise and study the intrinsic spectral evolution at short time-scales. The Poisson noise also broadens significantly the distribution of Γ (Fig. 15). However, we need to be cautious to claim that the broadening of the Γ distribution from the simulation (as shown in the right-hand column in the Fig. 15) is completely caused by the intrinsic short-time-scale spectral fluctuation. It might also be introduced by our algorithm, which we investigate next.

In order to test the effects of the algorithm on the recovery of Γ values, we produced a new synthetic data set by using all information described above with the exception of the Γ values, which were fixed at the single value of 1.69. In other words, we produced a set of light curves with null time/phase lags, which were therefore identical except for their normalizations. We added Poisson noise and studied the output Γ distribution after applying our algorithm. The histograms of Γ for the zero-lag synthetic data, the values obtained through direct spectral fitting and those from our reconstruction are plotted in Fig. 16. From our test, we can derive that the systematic broadening of the Γ distribution introduced by our procedure is very limited (less than 0.004 or 0.24 per cent) (note the horizon scale of the right-hand panel in Fig. 16).

The observed Γ distribution shown in the left-hand column of Fig. 15 is the result of both the intrinsic distribution in Γ and the broadening due to noise. Thus, the standard deviation σ_Γ of the intrinsic spread in Γ can be quantitatively estimated as

$$\sigma_\Gamma = \sqrt{\sigma^2_{\rm data} - \langle e^2 \rangle},$$

where $\sigma^2_{\rm data}$ is the variance of the Γ distribution observed from real data (left-hand column of Fig. 15) and $\langle e^2 \rangle$ is the mean square error on Γ obtained from the spectral fitting analysis. For the shortest time-scale of 0.0625 s, we have obtained 25 600 values of Γ by fitting the dynamic energy spectra, obtaining $\sigma^2_{\rm data} = 0.0141$, $\langle e^2 \rangle = 0.0113$. $\sigma^2_{\rm data}$ is approximately equal to $\langle e^2 \rangle$, which proves that the spread in Γ at a short time-scale from the real data is mostly due to the Poisson noise. With the above equation we obtain $\sigma_\Gamma = 0.0529$, close to the standard deviation of 0.0517 calculated from the simulation data (the top-right panel in Fig. 15). For the longer time-scales, the two values are also found to be comparable. Again, this fact supports that our algorithm is capable of filtering the Poisson noise and reveal the intrinsic distribution of Γ. It is interesting to note that Kotov, Churazov & Gilfanov (2001) assumed the small variations of the power-law index in their time-lag model. Our study on the intrinsic Γ distribution thereby gives an evidence for their assumption.

We can therefore draw three conclusions as follows.

(i) The broadening of Γ distribution introduced by our algorithm can be neglected and the histograms in the right-hand column of

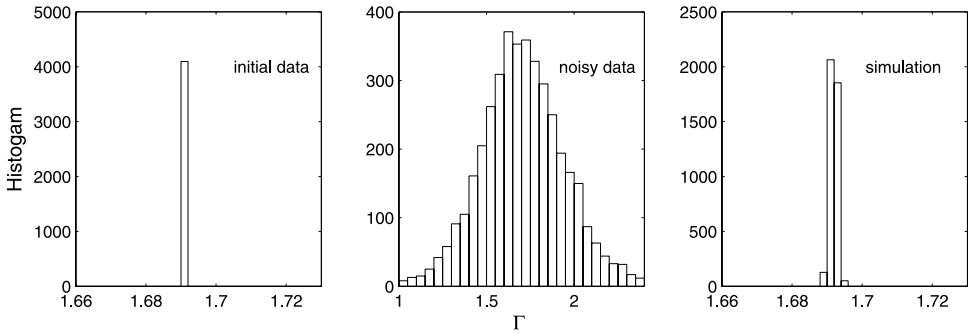

Figure 16. The histograms of the power-law photon index Γ at 0.0625 s for data with Γ distributed as a δ function at 1.69 (left), the same data with added Poisson noise (middle) and the reconstructed data with our synthetic algorithm (right).

Fig. 15 reflect the intrinsic spectral variation of the source on these time-scales.

(ii) Our synthetic method is indeed powerful in filtering Poisson noise and recovering underlying statistical properties of the source data which are masked by Poisson noise.

(iii) The Γ distribution obtained through short-time spectral fitting of the data is completely dominated by the effects of Poisson noise and cannot be used to ascertain the real distribution.

A different approach studied by Revnivtsev, Gilfanov & Churazov (1999) was followed through Fourier-resolved spectra, which give the energy-dependent variability amplitude in a certain frequency range. The frequency-dependent spectral variability revealed by the Fourier-resolved spectra, however, cannot be immediately linked with the variation of spectral indices studied here. First, the time-scale in our work refers to the time bin size other than the reciprocal of frequency. The variations sampled on short time bins come from both low-frequency and high-frequency variabilities presented in PDS. However, the power density or variability calculated by binning the time series is comparable to that on the corresponding frequency (see Wu et al. 2009, and references therein), and the time-scale can be taken in practice as the time bin size. What is more important is that although Fourier-resolved spectra have a form similar to energy spectra, they have a totally different physical interpretation. Therefore, the comparison of their spectral indices is not useful. For example, the Fourier-resolved spectrum was found to be harder at higher frequencies, which only suggests that the hard X-ray radiation component exhibits larger variation amplitude compared with the soft radiation as the frequency increases. The only possible connection we can seek between this phenomenon and our study on the energy spectral indices is that the variations of Γ on a short time-scale are probably mainly due to the hard spectral component.

3.5 Extrapolation and interpolation

The synthetic algorithm can be used to produce the data with better time and energy resolution than the real data, after some additional hypotheses are made. A two-Lorentzian model has been used to describe the PDS and lag spectra (see Sections 2.2 and 2.3). If we extrapolate the model frequency beyond the Nyquist frequency of the original data, we are able to derive a synthetic light curve with higher time resolution. Moreover, we have derived the energy dependence of the best-fitting parameters for the PDS and lag spectra. By interpolating these functions, we can obtain the PDS and lag spectra (and therefore produce synthetic data) on an energy grid finer than the initial energy resolution. The hypothesis on which the extrapolation and interpolation are based is that PDS and time lag evolve smoothly in frequency and energy and that they can be extrapolated from the observed values.

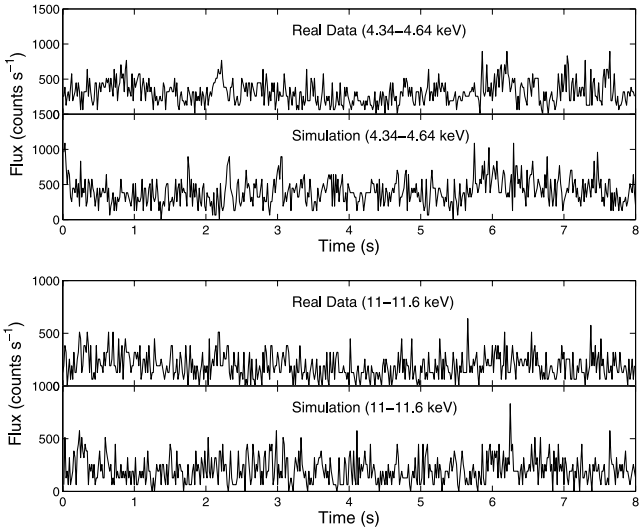

Figure 18. The real and simulated light curves in two narrow energy bins of 4.34–4.64 keV (top panel) and 11–11.6 keV (bottom panel). The simulation is based on the interpolation of the best-fitting parameters for PDS and time-lag spectra between energy bins.

The results of the extrapolation to higher frequency and of the interpolation between energies are shown in Figs 17 and 18, respectively. The time resolutions in the two panels of Fig. 17 are 0.004 and 0.001 s, respectively, smaller than the time resolution of 0.015 625 s for the real data. The dynamic energy spectra were also studied on these time-scales and the resulting power-law photon indices are plotted in the figure. Poisson noise is *not* added to the light curve, and an explicit positive correlation between the photon index and flux can be observed. The best-fitting parameters of PDS and lag spectra were interpolated to two additional narrow energy bins, 4.34–4.64 and 11–11.6 keV, respectively (see open circles in Figs 5 and 7). As mentioned, the original data have 64 energy bins, which we rebinned into eight coarse energy bins as the input of our algorithm. This allowed us to compare the simulation and the real data in these two narrow energy bins, as shown in Fig. 18.

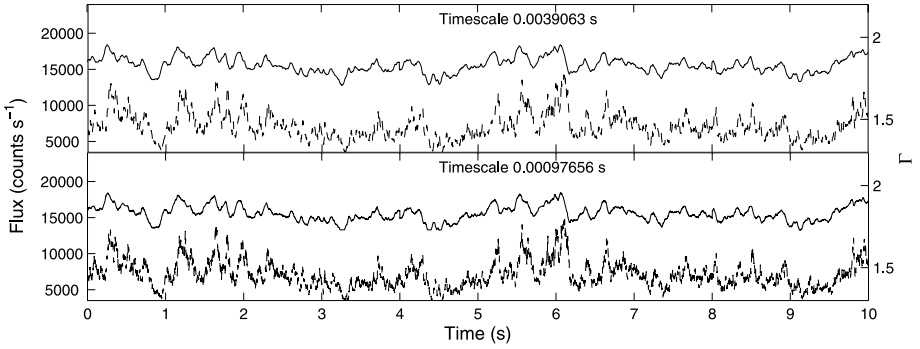

Figure 17. The time evolution of the power-law photon index Γ (solid line) and 0.14–25 keV net flux (dashed line) derived from the simulation based on the extrapolation of PDS and time-lag spectra to higher frequency. The time-scales of 0.004 s (top) and 0.001 s (bottom) are shorter than the time resolution of the original data.

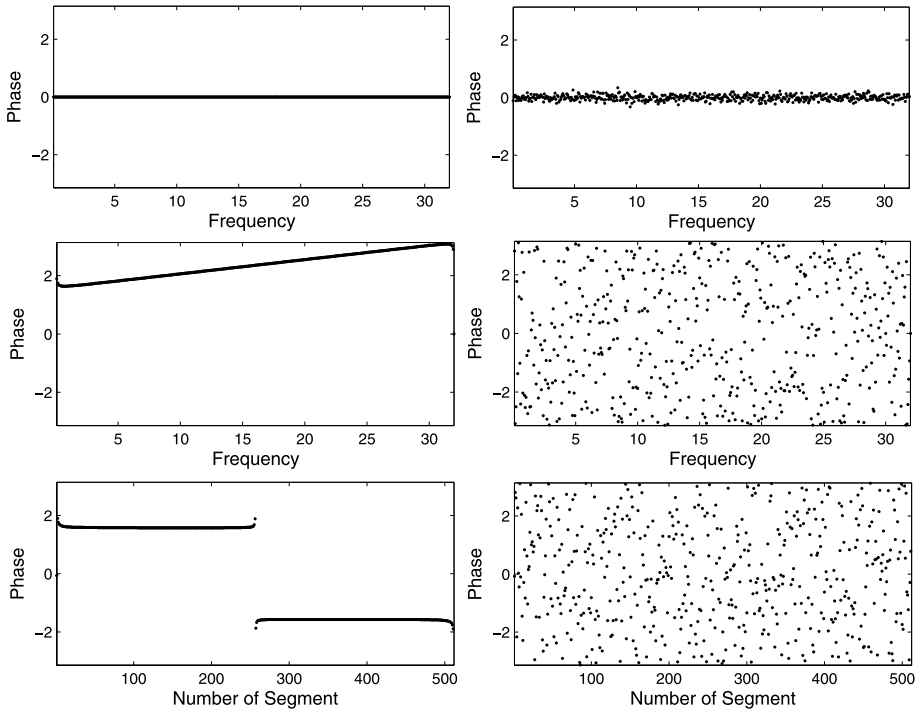

Figure 19. Top: the phases of the initial light curve with (left) and without (right) additional Poisson noise. Middle: the phases at different frequencies for a single segment after arbitrarily choosing the starting point and segment (left) and the case with additional Poisson noise (right). Bottom: the phase for different segments for a fixed frequency after arbitrarily choosing the starting point and segment (left) and the case with additional Poisson noise (right).

3.6 Phase and noise

In Section 3.3, we wondered whether an exponentiation is necessary after the IFT to produce a light curve similar to the real one when random phases are assumed. This is peculiar since we can strictly recover the time series with the IFT alone, if we know the phases. It is therefore logical to deduce that the phases of the real data cannot be random. The exponential transformation is simply a compensation for the incorrect assumption of random phases. The effect of the exponential transformation in the time domain should be represented in the Fourier domain as a modification of phase, since there is practically no effect on the PDS shape. The fact that the phase cannot be totally random and independent can also be revealed by higher order variability properties, e.g. the bicoherence, a higher order statistics measuring the degree of coupling between variations on different time-scales. A non-zero bicoherence indicates that there exist correlations between the phases at different frequencies within a single energy band (Maccarone & Coppi 2002; Uttley et al. 2005). Hence, the phases cannot be independent nor can they be strictly uncorrelated. From Fig. 9, the phases appear to be uniformly distributed over the range $(-\pi, \pi]$ and a correlation between them is not apparent. A higher order statistical test such as bicoherence would probably show these effects. The assumption about the phase distribution presented in Section 3.3 can be considered as a lower order approximation and is appropriate for the practical purpose of simulating the linear time series which will later be exponentially transformed.

Moreover, it is possible that the procedure of segmenting the data conceals the underlying phase. In order to investigate this possibility, we performed this experiment (see Fig. 19). With an arbitrary PDS and zero lag between energies, we synthesized an initial light curve. Its time lag is of course zero at all frequencies. If we add Poisson noise to the light curve, the phases [calculated from the fast Fourier transform (FFT)] would be scattered around zero. In practice, the start time of the observation is arbitrary and the long light curve is split into short segments to calculate the phase. We chose an arbitrary starting point and calculated the FFT in 1024-s long segments. The phases at different frequencies for a single segment and the phases at different segments for a single frequency apparently deviate from zero. If we again add Poisson noise, the phase becomes totally random. Therefore, even if the intrinsic phase is not random or uncorrelated, the arbitrary selection of a starting point and the presence of additional Poisson noise would make the detected phase appear random. The underlying phase is likely not random, although this cannot be inferred directly by the data. Up to now, we still know little about the intrinsic distribution of the phases and cannot propose a hypothesis more reasonable than the random distribution. Therefore, we stick to the assumption of random phases uniformly distributed between $(-\pi, \pi]$ throughout this work.

4 CONCLUSIONS

We have developed an algorithm to produce synthetic light curves from the observed properties of Cyg X-1. The algorithm is based on the previously established time series simulation method (e.g. Timmer & König 1995; Uttley et al. 2005) and improved in three aspects as follows: (i) besides the PDS, information from additional measurements such as energy spectra and lag spectra is used as input; (ii) the synthetic time series is required to reproduce almost all the timing and spectral properties; and (iii) the possible applications of the method are explored, including filtering Poisson noise and data extrapolation. It is model independent, unlike the attempts to restore the timing and spectral properties through a physically interesting model and parameters (e.g. Arévalo & Uttley 2006). The simulation does not provide more information than in the original

PDS, lag spectra, etc. What we do is to allow a different view of the same data.

By combining all known information about the observed variability, a reasonable assumption on the distribution of phases and prior knowledge about the Poisson noise power, we can obtain synthetic data which are not affected by Poisson noise. From these synthetic data, we can explore the spectral variability of the source on short time-scales, where the real data are noise-dominated. We showed that the observed distribution of spectral indices of Cyg X-1 on short time-scales is completely dominated by Poisson effects, since even simulated data with a δ distribution in spectral indices yields the same output distribution. From the output of our algorithm, we have recovered the real underlying distribution, under a relatively small number of assumptions. Our method shows that our current data are sufficient to reproduce the observed properties with good accuracy. Future missions will yield much higher statistics and will allow the exploration of spectral variability at higher frequencies and with better spectral resolution.

ACKNOWLEDGMENTS

We thank Phil Uttley for providing us with the code for the method published in Uttley et al. (2005). YXW thanks T. P. Li and S. N. Zhang for useful comments. We also appreciate the anonymous referee for the very insightful suggestions, which help improving this paper a lot. This work was supported by contract PRIN-INAF 2006.

REFERENCES

Arévalo P., Uttley P., 2006, MNRAS, 367, 801
Belloni T., Hasinger G., 1990, A&A, 227, L33
Belloni T., van der Klis M., Lewin W. H. G., van Paradijs J., Dotani T., Mitsuda K., Miyamoto S., 1997, A&A, 322, 857
Cui W., Zhang S. N., Focke W., Swank J. H., 1997, ApJ, 484, 383
Cui W., Feng Y.-X., Ertmer M., 2002, ApJ, 564, L77
Davies R. B., Harte D. S., 1987, Biometrika, 74, 95
Davis B. M., Hagan R., Borgman L. E., 1981, Comput. Geosci., 7, 199
Done C., Madejski G. M., Mushotzky R. F., Turner T. J., Koyama K., Kunieda H., 1992, ApJ, 400, 138
Gierliński M., Zdziarski A. A., 2003, MNRAS, 343, L84
Johnson G. E., 1994, Proc. IEEE, 82, 270
König M., Timmer J., 1997, A&AS, 124, 589
Kotov O., Churazov E., Gilfanov M., 2001, MNRAS, 327, 799
Leahy D. A., Darbro W., Elsner R. F., Weisskopf M. C., Kahn S., Sutherland P. G., Grindlay J. E., 1983, ApJ, 266, 160
Liu C. Z., Li T. P., 2004, ApJ, 611, 1084
Liu B., Munson D. Jr, 1982, IEEE Trans. Acoustics, Speech Signal Processing, 30, 973
Lochner J. C., Swank J. H., Szymkowiak A. E., 1991, ApJ, 376, 295
Maccarone T. J., Coppi P. S., 2002, MNRAS, 336, 817
Miyamoto S., Kitamoto S., Mitsuda K., Dotani T., 1988, Nat, 336, 450
Miyamoto S., Kitamoto S., Ida S., K., Negoro H., Terada K., 1992, ApJ, 391, L21
Negoro H., Miyamoto S., Kitamoto S., 1994, ApJ, 423, L127
Nowak M. A., Vaughan B. A., Wilms J., Dove J. B., Begelman M. C., 1999, ApJ, 510, 874
Pottschmidt K., Koenig M., Wilms J., Staubert R., 1998, A&A, 334, 201
Pottschmidt K. et al., 2003, A&A, 407, 1039
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in C, 2nd edn. Cambridge Univ. Press, Cambridge
Revnivtsev M., Gilfanov M., Churazov E., 1999, A&A, 347, L23
Terrell N. J. J., 1972, ApJ, 174, L35
Timmer J., König M., 1995, A&A, 300, 707
Uttley P., McHardy I. M., Vaughan S., 2005, MNRAS, 359, 345
van der Klis M., 2006, in Lewin W. H. G., van der Klis M., eds, Compact Stellar X-ray Sources. Cambridge Univ. Press, Cambridge
Vaughan B. A., Nowak M. A., 1997, ApJ, 474, L43
Vaughan S., Edelson R., Warwick R. S., Uttley P., 2003, MNRAS, 345, 1271
Wilms J., Nowak M. A., Pottschmidt K., Pooley G. G., Fritz S., 2006, A&A, 447, 245
Wu Y. X., Li T. P., Belloni T. M., Wang T. S., Liu H., 2009, ApJ, 695, 921
Zdziarski A. A., Poutanen J., Paciesas W. S., Wen L., 2002, ApJ, 578, 357
Zhang W., Jahoda K., Swank J. H., Morgan E. H., Giles A. B., 1995, ApJ, 449, 930
Zoghbi A., Fabian A. C., Uttley P., Miniutti G., Gallo L. C., Reynolds C. S., Miller J. M., Ponti G., 2010, MNRAS, 401, 2419

This paper has been typeset from a TEX/LATEX file prepared by the author.