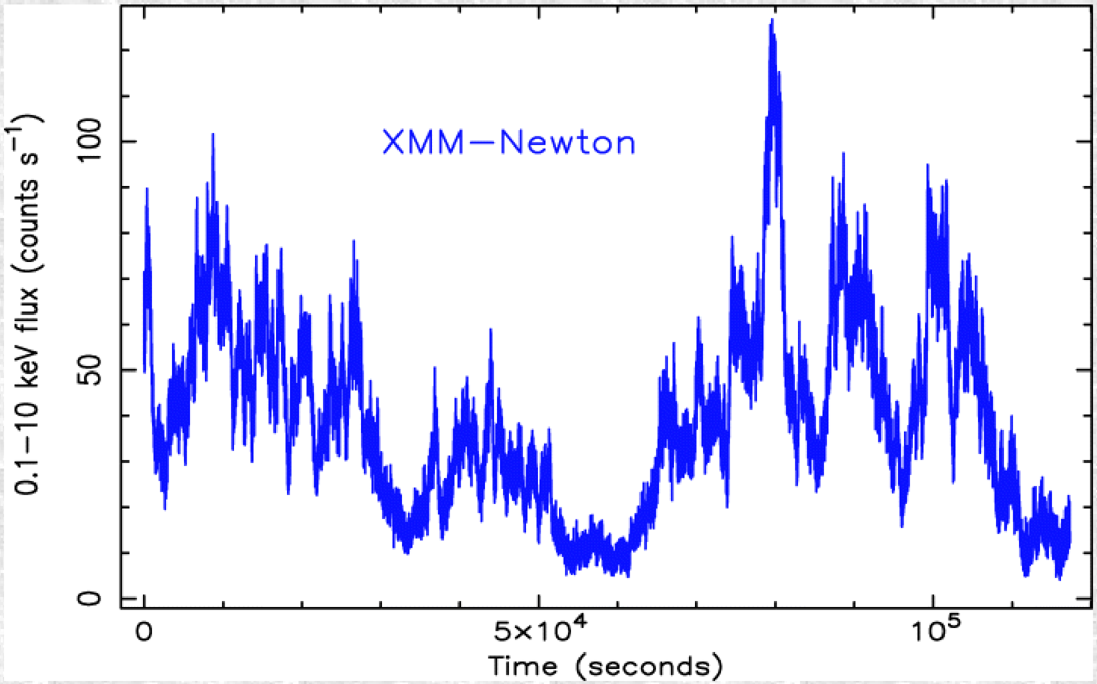
Fourier analysis of AGN light curves: practicalities, problems and solutions

Iossif Papadakis

McHardy et al., 2004, MNRAS, 348, 783

Study of the X-ray variability of NGC4051, using long-term RXTE data and short-term XMM data.



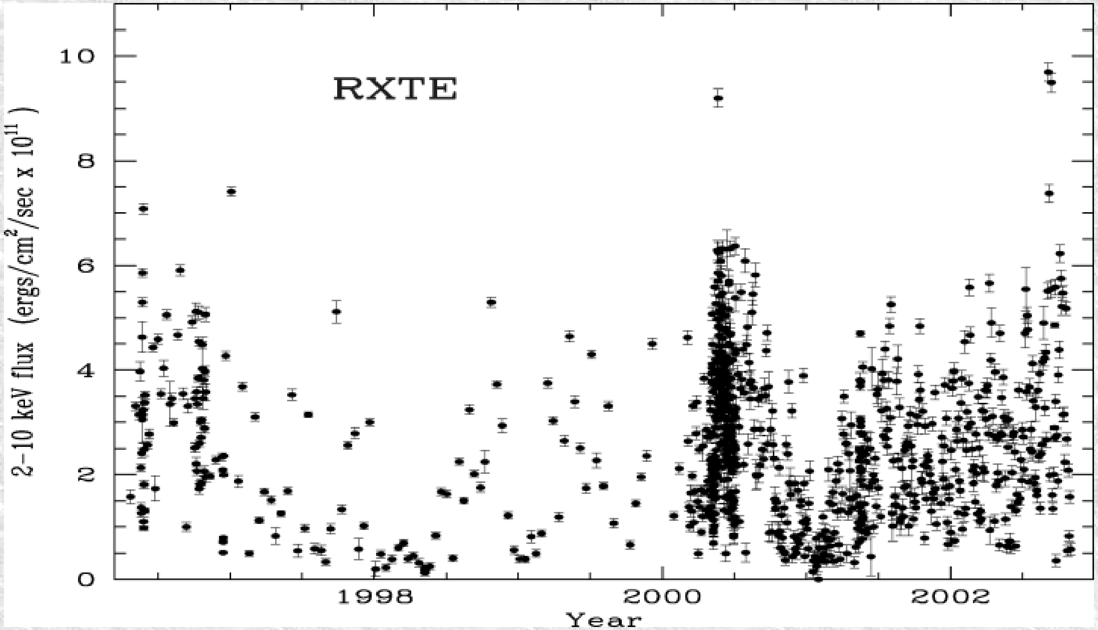
117 ksec long*, 23400 points (lc binned in 5 sec).

Can be used to estimate the PSD at frequencies: 10^{-5} – 0.1 Hz

*could correspond to ~ 1.17 sec for a 10 solar mass BH...

480 obs, 1 ksec long each, with RXTE, performed over a period of ~ 6 years!

Can be used to estimate the PSD at frequencies: $5x10^{-9}$ – $5x10^{-5}$ Hz.



???!?!?!?!

Bibliography:

1) M. B. Priestley, “Spectral Analysis and Time Series, Vol I. And II Academic Press.

2) T. J. Deeming, 1975, Astrophysics and Space Science, 36, 137

3) Scargle J. D., 1982, ApJ, 263, 835

4) Uttley P., et al., 2002, MNRAS, 332, 231

4) Timmer J. & Koenig M., 1995, A&A, 300, 707

1) Introduction (and revision)

Suppose we want to study the properties* of a continuous, stochastic/random process, $X(t)$.

In most cases, the most interesting characteristic of such a random process is the presence of a 'memory' in the system: for some reason, the value of the process at time t, $x(t)$, is 'influenced' by the state of the system at t-τ, $x(t-\tau)$. This dependence/memory is usually quantified by the the 'auto-covariance' function:

$$R(\tau)=\langle[x(t)-\overline{x}][x(t-\tau)-\overline{x}]\rangle$$

*In my case, the word "properties" means the 'power spectral density function' of the process.

The Fourier transform of R(τ) (which can, in principle, reveal `characteristic frequencies', associated with the 'memory' of the system)

$$=\int_{-\infty}^{+\infty} R(\tau)\, e^{-i2\pi\nu\tau}\, d\tau = 2\int_{-\infty}^{+\infty} R(\tau)\cos(2\pi\nu\tau)\, d\tau \;=$$

'power spectral density function' of the process, *P(v)* (*P(v)dv*= contribution to the total variance of components in X(t) with frequencies between *v* and *dv*). 'Normally',

a) we observe a process (i.e. the X-ray emission from a GBH binary, an AGN, etc), *N* times, and we get:

x_i at $t_i=t_0+i\Delta t$, $i=1,2,\ldots,N$

b) we define *N/2* frequencies,

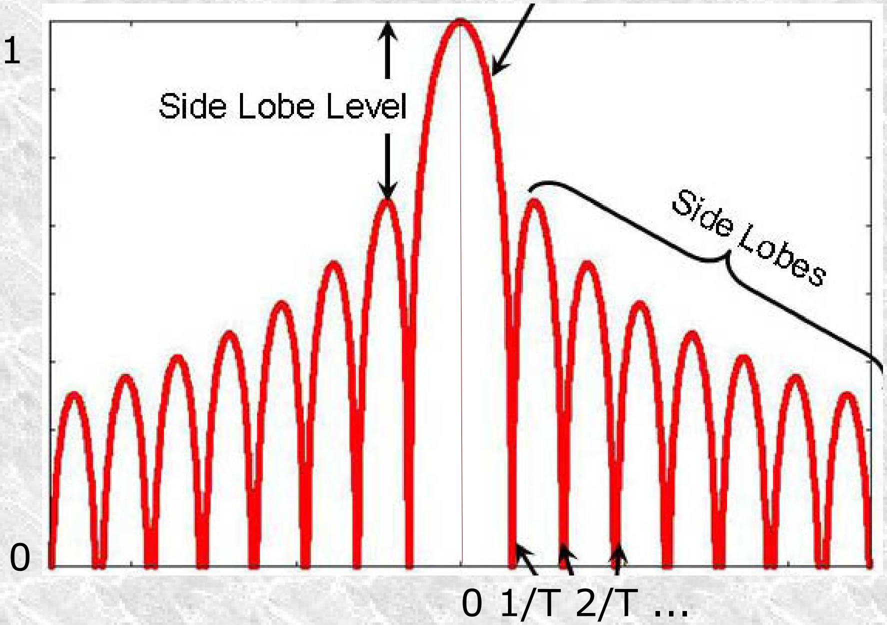
$\nu_i=i/(N\Delta t)$, $i=1,2,..,N/2$ **(1)**

c) at which we compute the discrete Fourier transform (or periodogram) of the observed time series:

$$I(\nu)=\frac{1}{N\Delta t}\left|\sum_{i=1}^{N} x_i e^{-i2\pi\nu t}\right|^2=\frac{1}{N\Delta t}\left\{\left[\sum_{i=1}^{N} x_i\cos(2\pi\nu t)\right]^2+\left[\sum_{i=1}^{N} x_i\sin(2\pi\nu t)\right]^2\right\} \quad \textbf{(2)}$$

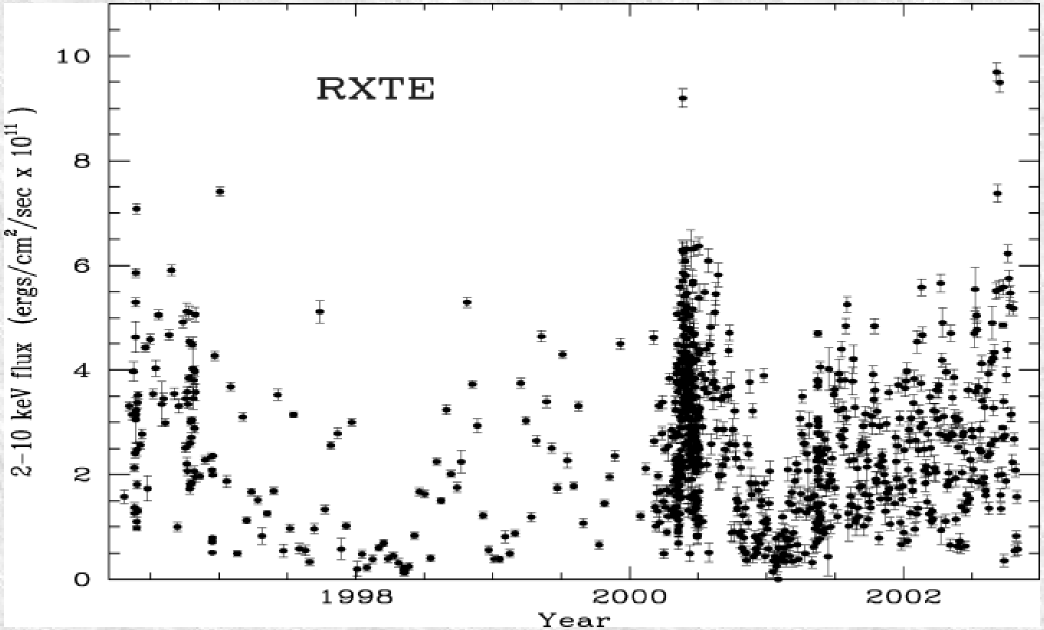
Relation between *P(v)* and *I(v)*: $\langle I(\nu)\rangle = \int_{\frac{-1}{2\Delta t}}^{\frac{+1}{2\Delta t}} P(\nu')\, F_N(\nu'-\nu)\, d\nu'$

$F_N(v'\text{-}v)$ (Fejer Kernel):



As $N \to \infty$, $F(\nu' - \nu) \to \delta(\nu)$. We therefore say *I(v)* is an unbiased estimate of *P(v)*.

2) Back to the 'real' world



How can we estimate P(v) from 'this' data set?

1st Problem: At which frequencies??? Can we still use the set of frequencies defined by eq. (1).

v_{min} is still well defined: $v_{min}=1/T$. But v_{max}?

The min time difference, Δt_{min}, between neighbouring points is 6 hrs, but it does not seem 'right' to accept $1/\Delta t_{min}$ as v_{max}. An 'average' Δt seems to be more appropriate in order to define v_{max}. But which mean should we use?

2nd Problem: Suppose we define a set of frequencies, how are we going to estimate *P(v)*?

We could still use eq. **(2)** (there is no physical law against using it!), but what will then be the relation between *P(v)* and *I(v)*? Well, even in this case,

$$\langle I(\nu)\rangle = \int_{\frac{-1}{2\Delta t_{min}}}^{\frac{+1}{2\Delta t_{min}}} P(\nu')\, V_N(\nu'-\nu)\, d\nu', \qquad \textbf{(3)}$$

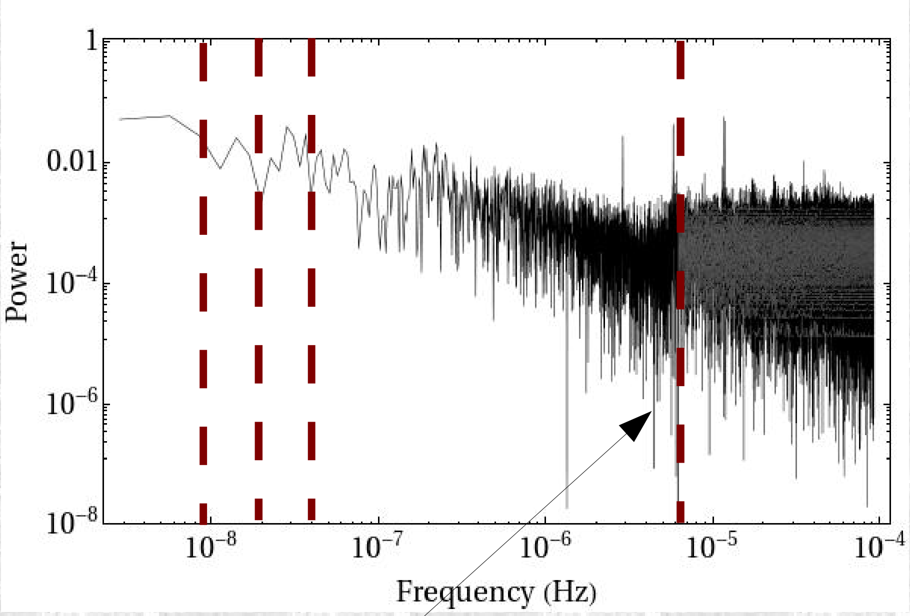
where: $V_N(\nu) = \left|\sum_{i=1}^{N} e^{i2\pi\nu t_i}\right|^2 = \left[\sum_{i=1}^{N} \cos(2\pi\nu t_i)\right]^2 + \left[\sum_{i=1}^{N} \sin(2\pi\nu t_i)\right]^2,$ **(4)**

is the 'window function', which obviously depends on the sampling pattern of the lightcurve (and cannot be approximated by a simple analytic function).

This is $V_N(\nu)$ for the most recent NGC4051 RXTE light curve (T=4000 days).

It shows 'minima' at 3/T, 6/T, 9/T,...

So, we could compute $I(\nu)$ at these frequencies, ... up to?



$\nu_{max} \sim 6\times10^{-6}$ Hz, above which $V_N(\nu)$ does not decrease anymore with increasing frequency.

Summary so far for the power-spectral estimation in the case of unevenly sampled light curves:

a) Use eq. **(4)** to compute the 'window function'

b) Check if there are 'local' minima in $V_N(v)$, and if they are roughly 'evenly' spaced. If yes,

c) Use as v_{min} the frequency at which the first minimum in the $V_N(v)$ appears, and define a set of frequencies: $v_i = i * v_{min}$, $i=1,2,...$ up to the point where $V_N(v)$ flattens.

d) Use eq. **(2)** to compute $I(v)$. Bin nearby values (in log-log) space to get your 'final' estimates, $PSD_{obs}(v)$.

... of $P(v)$. However, equation **(3)** still holds, and the distortion of $I(v)$ due to the convolution of $P(v)$ with the window function can be severe...

3) Modeling of the sampled PSD

(yes, everything is possible...)

You can learn something about the timing properties of the process by fitting a model to the observed PSD. So, one could proceed as follows:

a) Assume a model PSD, say $P_{mod}(v)$,

b) Solve equation (3) to compute the predicted $I_{mod}(v_i)$,

c) Bin $I_{mod}(v_i)$, exactly as done with the observed $I(v_i)$'s, to compute $PSD_{mod}(v)$,

d) compare $PSD_{mod}(v)$ with $PSD_{obs}(v)$ (using for example the Levenberg-Marquardt χ^2 minimization method).

HOWEVER, ...

... contrary to the evenly sampled case,

a) the distribution of $I(v_i)$'s is not known apriori, and even if you bin the $I(v_i)$'s, you cannot tell if the distribution of $PSD_{obs}(v)$ will be well approximated by a Gaussian

b) you are not certain about the errors of $PSD_{obs}(v)$, and

c) the $I(v_i)$'s (and $PSD_{obs}(v)$'s) are *not* independent random variables.

All the above imply that, if you apply a χ^2 minimization technique to compare $PSD_{mod}(v)$ with $PSD_{obs}(v)$,

a) you cannot tell whether the model is a 'good' one (i.e. you cannot judge the 'goodness' of the model fit),

b) you cannot estimate confidence limits on the best-fit model parameter values.

So, what is the solution???

4) Monte carlo simulations...

Here is a rough guide of want you may want to do, if you want to fit a model to the PSD of a light curve like the one I showed you before for NGC4051.

a) Suppose you want to consider a model of the form:

$$P_{mod}(\nu)=A(\nu/\nu_{br})^{-\alpha},\ \nu>\nu_{br}$$
$$P_{mod}(\nu)=A(\nu/\nu_{br})^{-\beta},\ \nu<\nu_{br}$$

b) And suppose you 'believe' that: $1<\alpha<2$, $0<\beta<1$, and 10^{-5} Hz $<\nu_{br}<10^{-3}$ Hz.

c) Consider all the 'possible' α, β and ν_{br} values within these parameter values (i.e. values of α and β in steps of 0.05, and values of ν_{br} in steps of 1.26×10^{-5} Hz). This implies <u>*8000*</u> combinations of all possible model parameter values.

For *each* combination of model parameters,

a) Produce at least 500 synthetic light curves, which will be evenly spaced, with a total duration 10×longer than T_{obs}, at least, and a $\Delta t = \Delta t_{min,obs}$ (at least).

b) Re-ample these light curves, in a way that will resemble the observed light curve.

c) Estimate the PSD, in the same way as you did for the observed light curve.

d) At each frequency, compute the mean 'model' PSD ($<PSD_{mod}>$), and the spread (σ_{mod}) of the 500 PSDs around $<PSD_{mod}>$, and

e) compute, for each synthetic PSD, a 'χ^2' value:

$$\chi^2 = \frac{\sum_{i=\nu_{min}}^{\nu_{max}} \left(PSD_{mod,i} - \langle PSD_{mod} \rangle\right)^2}{\sigma^2_{mod,i}} \quad \textbf{(5)}$$

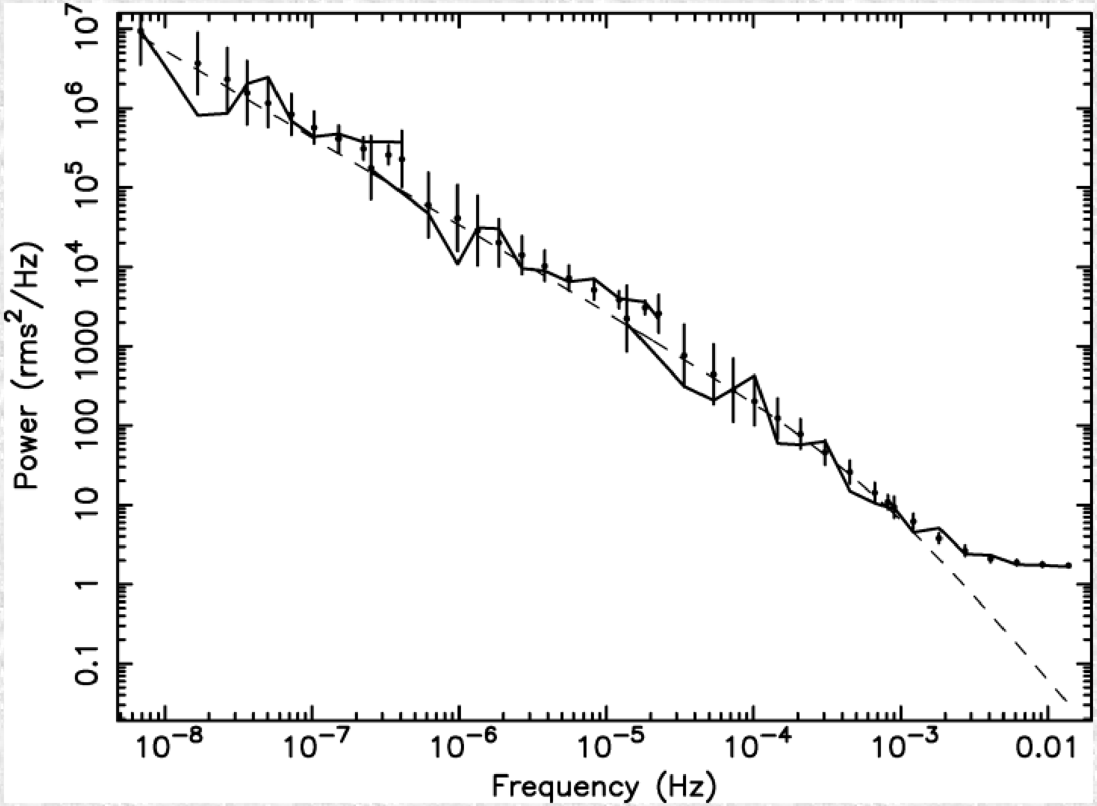
Go back to $PSD_{obs}(v)$.

a) For <u>each</u> model parameter combination (these 8000 possible variations...), use eq. (5) to estimate χ^2_{obs}.

b) Find that particular 'model parameter values combination' that will give you the minimum χ^2_{obs}. This is your <u>'best fit'</u> model!

c) <u>Is it a 'good' one?</u> Use the distribution of the χ^2_{mod} values for 'this' model, to judge if the fit is 'acceptable' (say if χ^2_{obs} is smaller say than at least 10% of all the χ^2_{mod} values).

d) It is more complicated to estimate confidence limits on the best-fit parameter values.



Here it is for NGC4051! The observed PSD is well fit by a power-law model which has a slope of ~ -1 at frequencies below ~ 7×10^{-4} Hz, and a slope of ~ -3 at higher frequencies!

IT CAN BE DONE!!!! I told you so...

It is a difficult, time consuming, tiring job, but it can give interesting results!

Thank you.