J. Astrophys. Astr. (2011) **32,** 79–86

Methods for the Quasi-Periodic Variability Analysis in Blazars

Y. Liu*, J. H. Fan, H. G. Wang & G. G. Deng
Center for Astrophysics, Guangzhou University, Guangzhou 510006, China.
**e-mail: pinux@gzhu.edu.cn*

Abstract. In this work, four methods are introduced to analyse the unevenly sampled data. The four methods are power spectral density, autocorrelation function, structure function and Jurkevich method. Some interesting mathematical links are derived amongst the four methods. These links show that the four methods have the same performance. For the Jurkevich method, the effect of the width of bins is apparent. If the width is half the time scale, the method has better performance.

Key words. Method: quasi-periodic analysis—unevenly sampled time series analysis.

1. Introduction

Most astronomical data are unevenly sampled in time, sometimes with large gaps which recur over a regular period on their own. In any case, unevenly sampled time series introduces myriad complications into the traditional analysis methods. Many attempts have been made to deal with these problems. Here, four methods to analyse the unevenly sampled data are introduced. These techniques are applied to the variability analysis in blazars in optical and radio bands, to search for possible quasi-periodic signals.

2. Power spectral density (PSD)

In statistical signal processing and physics, the power spectral density (PSD) is a positive real function of a frequency variable associated with a stationary stochastic process. Intuitively, the spectral density captures the frequency content of a stochastic process and helps identify periodicity (Fig. 1).

Define a continuous function $y(t), t \in [a, b]$. Let $[\tau_1, \tau_2] \subset [a, b]$. The timespan $s = \tau_2 - \tau_1$ of y is referred to as 'lag'. Then the mean of y on $[\tau_1, \tau_2]$ is:

$$\langle y \rangle_t = \frac{1}{\tau_2 - \tau_1} \int_{\tau_1}^{\tau_2} y(t)\, \mathrm{d}t. \tag{1}$$

Define

$$Y(t) = y(t) - \langle y \rangle_t .$$

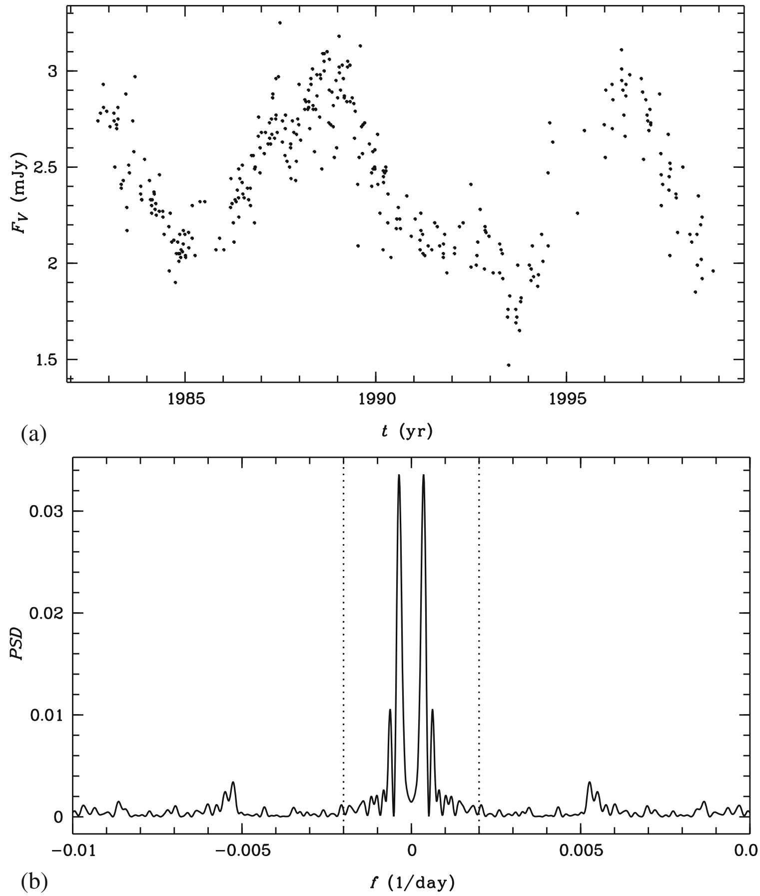


Figure 1. **(a)** The light curve of 0605-085 at 4.8 GHz. **(b)** The corresponding PSD, $f_0 = \pm 0.002$ day^{-1} are marked with dashed vertical line.

For the continuous case, the Fourier transform of $Y(t)$ on $[\tau_1, \tau_2]$ is:

$$F_w(f) = \frac{1}{\tau_2 - \tau_1} \int_{-\tau_1}^{\tau_2} Y(t) \mathrm{e}^{-j2\pi f t} \, \mathrm{d}\, t = \left\langle Y(t) \mathrm{e}^{-j2\pi f t} \right\rangle_t .$$

Then the PSD(f) of y on $[\tau_1, \tau_2]$ can be defined as:

$$\begin{aligned} \mathrm{PSD}_w(f) &= F_w(f) \cdot \tilde{F}_w(f) \\ &= \frac{1}{(\tau_2 - \tau_1)^2} \int_{\tau_1}^{\tau_2} \int_{\tau_1}^{\tau_2} Y(t) Y(t') \cos[2\pi f (t - t')] \, \mathrm{d}\, t \, \mathrm{d}\, t' \end{aligned} \qquad (2)$$

Partition the interval $[\tau_1, \tau_2]$ into n divisions with $\tau_1 = t_0 < t_1 < \cdots < t_n = \tau_2$. Then $y_i = y(t_i)$ is an unevenly sampled time series. The mean of y_i can be defined as:

$$\langle y \rangle = \frac{1}{\tau_2 - \tau_1} \sum_{i=1}^{n} y_i \times (t_i - t_{i-1}).$$

Define

$$Y_i = y_i - \langle y \rangle ,$$

$$\mathrm{PSD}_w(f) = \frac{1}{(\tau_2 - \tau_1)^2} \sum_{i=1}^{n} \sum_{j=1}^{n} Y_i Y_j \cos\left[2\pi f \left(t_i - t_j\right)\right] \times (t_i - t_{i-1}) \left(t_j - t_{j-1}\right) . \quad (3)$$

3. Auto-correlation function (ACF)

Auto-correlation is the cross-correlation of a signal with itself. The auto-correlation function (ACF) method, which was described in detail by Edelson & Krolik (1988) (also see Fan *et al.* 1998) is intended for the analysis of cross-correlation of two datasets.

This method can indicate the correlation of two variable time series with a time delay τ, and can be applied to the periodicity analysis of a unique temporal dataset. If there is a period, P in the light curve, then the ACF should show clearly whether the dataset is correlated by itself with time delays of $\tau = 0$ and $\tau = n \times P (n \in N)$ (Fan *et al.* 1998).

The classic ACF is performed on a discrete time series dataset $\{y_i : i = 1, 2, \ldots, n\}$ with time t_i uniformly-spaced, which is however not suitable for other time series that are not equidistantly sampled, such as most astronomical data. We thus generalize the ACF from equidistant sampling to uneven sampling (Fig. 2).

3.1 *Weighted average ACF*

The ACF(τ) ($\tau \geq 0$) of y on $[\tau_1, \tau_2]$ can be defined as:

$$\mathrm{ACF}(\tau) = \frac{\tau_2 - \tau_1}{\tau_2 - \tau_1 - |\tau|} \frac{\int_{\tau_1}^{\tau_2} \int_{\tau_1}^{\tau_2} Y(t)\delta(t' - t - \tau)Y(t')\,\mathrm{d}t\mathrm{d}t'}{\int_{\tau_1}^{\tau_2} Y^2(t)\mathrm{d}t} .$$

$\delta(t)$ is the Dirac delta function.

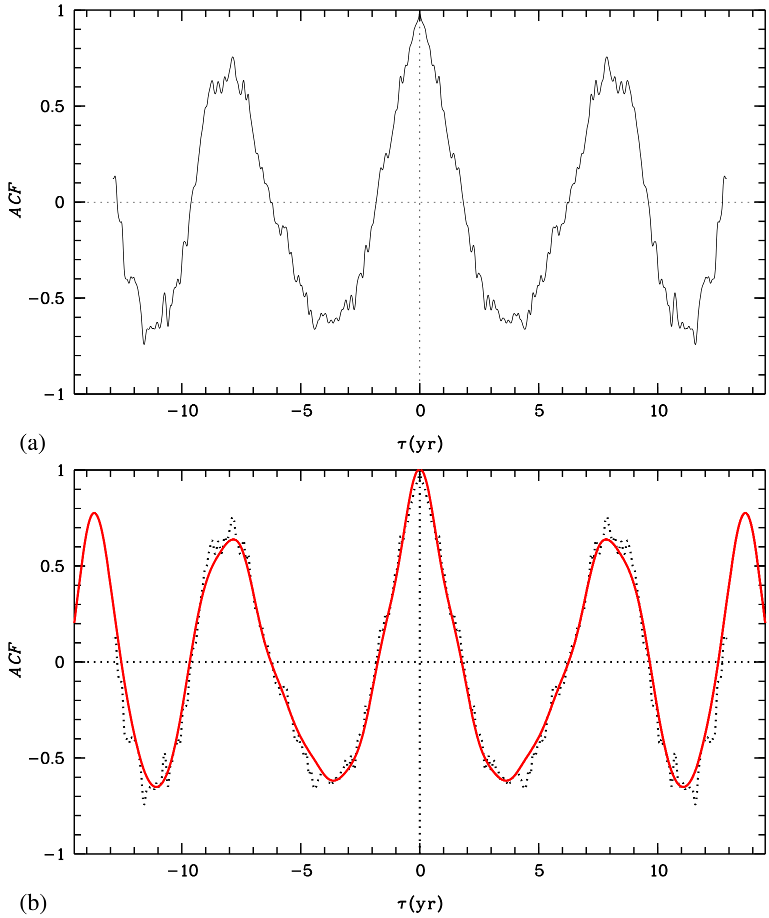
The Gauss function $w(t, \sigma)$ can be used instead of the $\delta(t)$ function,

$$w(t, \sigma) = \mathrm{e}^{-\frac{t^2}{2\sigma^2}} .$$

The ACF of unevenly sampled y_i can be defined as:

$$\mathrm{ACF}(\tau, \sigma) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} Y_i w(t_j - t_i - \tau, \sigma) Y_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} Y_i w(t_j - t_i, \sigma) Y_j} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w(t_j - t_i, \sigma)}{\sum_{i=1}^{n} \sum_{j=1}^{n} w(t_j - t_i - \tau, \sigma)} .$$

The question now is how to choose the value of σ. It is a dilemma, if giving a smaller σ, you will obtain a higher resolution in time domain and a more rough ACF



(a) (b)

Figure 2. **(a)** The ACF, the curve is ACF(τ, σ) with $\sigma = 16.97$ days. **(b)** The ACFs, the solid curve is ACF$(\tau,\ f_0)$, the dashed curve is ACF(τ, σ) in the bottom of figure **(a)**.

curve. According to our experience, for an approximate uniform sampled time series, $\sigma = s/n$ can be adopted.

3.2 *Obtain ACF from PSD*

The Wiener–Khinchin theorem relates the ACF to PSD. The theorem states that the PSD of a wide-sense-stationary random process is the Fourier transform of the corresponding ACF.

For a continuous case, the Fourier transform of $Y(t)$ is

$$F(f) = \int_{-\infty}^{\infty} Y(t) \mathrm{e}^{-j2\pi f t}\, \mathrm{d}t.$$

The PSD(f) is

$$\mathrm{PSD}(f) = F(f) \cdot \tilde{F}(f).$$

The ACF(τ) can be obtained as follows:

$$\mathrm{ACF}(\tau) = \frac{\int_{-\infty}^{\infty} \mathrm{PSD}(f) \mathrm{e}^{j2\pi f\tau}\,\mathrm{d}\,f}{\int_{-\infty}^{\infty} \mathrm{PSD}(f)\,\mathrm{d}\,f},$$

where $j = \sqrt{-1}$.

For the interval $[\tau_1, \tau_2]$, the windowing effects exist as:

$$\mathrm{ACF}(\tau,\, f_0) = \frac{\tau_2 - \tau_1}{\tau_2 - \tau_1 - |\tau|} \frac{\int_{-f_0}^{f_0} \mathrm{PSD}_w(f) \mathrm{e}^{j2\pi f\tau}\,\mathrm{d}\,f}{\int_{-f_0}^{f_0} \mathrm{PSD}_w(f)\,\mathrm{d}\,f}.$$

The factor $\frac{\tau_2-\tau_1}{\tau_2-\tau_1-|\tau|}$ is used to correct the windowing effects.

How to choose the value of f_0? You could get it from the PSD. The key point is to locate the main part of PSD within the interval $[-f_0,\, f_0]$.

4. Structure function (SF)

The first order structure function (SF) is a powerful tool to search for periodicities and time scales in a time series (see e.g., Rutman 1978; Simonetti *et al.* 1985; Paltani *et al.* 1997). Based on the basic idea used in ACF, we can also generalize the SF from the equidistant sampling to uneven sampling (Fig. 3).

4.1 *Weighted average SF*

Using the δ function, the SF of the continuous function $Y(t)$ can be defined as:

$$\mathrm{SF}(\tau) = \frac{\tau_2 - \tau_1}{\tau_2 - \tau_1 - |\tau|} \frac{\int_{\tau_1}^{\tau_2} \int_{\tau_1}^{\tau_2} \delta(t' - t - \tau)[Y(t) - Y(t')]^2\,\mathrm{d}\,t'\,\mathrm{d}\,t}{2\int_{\tau_1}^{\tau_2} Y^2(t)\mathrm{d}\,t}.$$

Using a Gaussian function $w(t, \sigma)$, we can define SF(τ, σ) of $Y_i = Y(t_i)$ as:

$$\mathrm{SF}(\tau, \sigma) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w(t_j - t_i - \tau, \sigma)(Y_i - Y_j)^2}{2\sum_{i=1}^{n} \sum_{j=1}^{n} Y_i w(t_j - t_i, \sigma) Y_j} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w(t_j - t_i, \sigma)}{\sum_{i=1}^{n} \sum_{j=1}^{n} w(t_j - t_i - \tau, \sigma)}.$$

As mentioned in ACF(τ, σ), $\sigma = s/n$ is good enough for some cases.

4.2 *Obtain SF from ACF*

There is also a direct relation between the SF and the ACF.

For a continuous case,

$$\mathrm{SF}(\tau) = \mathrm{ACF}(0) - \mathrm{ACF}(\tau) = \frac{\int_{-\infty}^{\infty} \mathrm{PSD}(f)[1 - \mathrm{e}^{j2\pi f\tau}]\,\mathrm{d}\,f}{\int_{-\infty}^{\infty} \mathrm{PSD}(f)\,\mathrm{d}\,f}.$$

For the interval $[\tau_1, \tau_2]$, the windowing effects exist, and we use the corrected ACF($\tau,\, f_0$) to obtain

$$\mathrm{SF}(\tau,\, f_0) = 1 - \mathrm{ACF}(\tau,\, f_0).$$

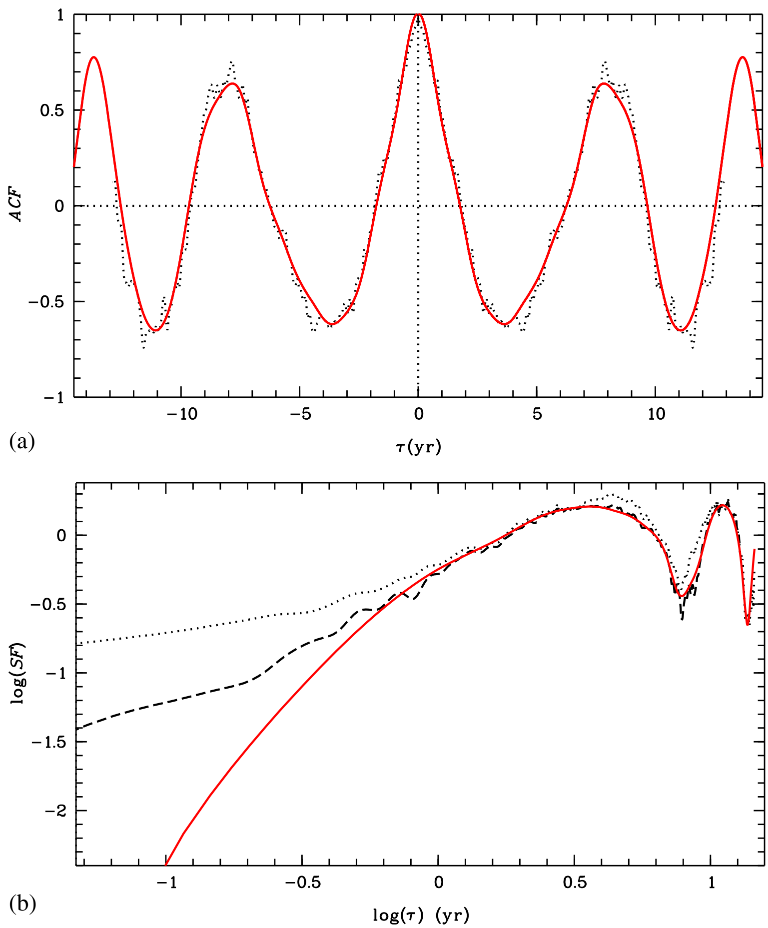


Figure 3. **(a)** The ACFs, the dashed curve is ACF(τ, σ) with $\sigma = 16.97$ days, the solid curve is ACF(τ, f_0). **(b)** The SFs, the dashed curve is SF(τ, σ) with $\sigma = 16.97$ days, the solid curve is SF(τ, f_0).

5. Jurkevich method (JV)

The Jurkevich method (Jurkevich 1971) is a computationally simple technique to detect the quasi-periods of an unevenly sampled time series, based on the expected mean square deviation. It tests a run of trial periods around which the data are folded.

$$\delta(t, \tau) = \begin{cases} 1, & \tau = 0, \\ \delta(\sin^2 \frac{\pi t}{\tau}), & \tau > 0, \end{cases}$$

$$S(t, \tau) = \frac{\int_{\tau_1}^{\tau_2} \delta(t - t', \tau) Y(t')\, \mathrm{d}t'}{\int_{\tau_1}^{\tau_2} \delta(t - t', \tau)\, \mathrm{d}t'},$$

$$\mathrm{JV}(\tau) = \frac{\int_{\tau_1}^{\tau_2} [Y(t) - S(t, \tau)]^2\, \mathrm{d}t}{\int_{\tau_1}^{\tau_2} Y^2(t)\, \mathrm{d}t}.$$

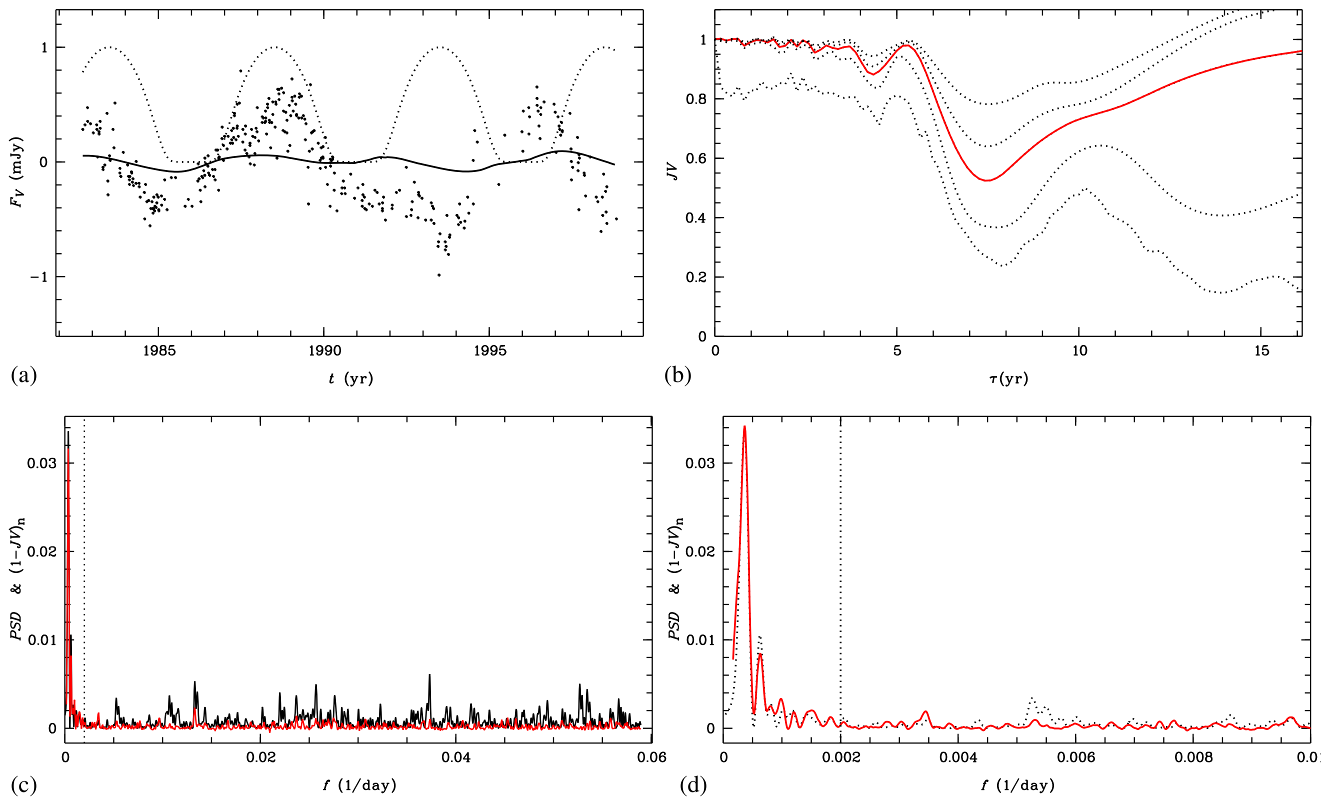


Figure 4. **(a)** The light curve of 0605-085 at 4.8 GHz, the dashed curve is $w(t, \tau, \lambda)$, the solid curve is $S(t, \tau, \lambda)$. **(b)** The corresponding JVs, the solid curve is the JV with $\lambda = 0.25$, the lower curve has a smaller λ. **(c)** The $1 - $ JV, the dashed curve is PSD. **(d)** The same but having a smaller time range.

Define

$$w(t,\tau,\lambda)=\begin{cases}1, & \tau=0,\\ e^{-\frac{\tan^2\frac{\pi t}{\tau}}{2\pi^2\lambda^2}}, & \tau>0,\end{cases}$$

$$S(t,\tau,\lambda)=\frac{\sum_{j=1}^{n} w(t_j-t,\tau,\lambda)Y_j}{\sum_{j=1}^{n} w(t_j-t,\tau,\lambda)},$$

$$\mathrm{JV}(\tau,\lambda)=\frac{\sum_{i=1}^{n}[Y_i-S(t_i,\tau,\lambda)]^2}{\sum_{i=1}^{n}Y_i^2}.$$

Here λ is an important parameter, the signal is more obvious with smaller λ, but it would introduce obvious false signals at times of τ. $\lambda = 0.25$ would be a good choice (Fig. 4).

6. Conclusion

We have demonstrated some interesting links between four methods to analyse the quasi-periodic variability. ACF is the Fourier transform of PSD, SF = 1 − ACF, JV and PSD have similar performance.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China (NSFC 10633010), the 973 Programme (2007CB815405) and the Bureau of Education of Guangzhou Municipality (No. 11 Sui-Jiao-Ke (2009)), Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (GDUPS) (2009), and Yangcheng Scholar Funded Scheme (10A027S).

References

Edelson, R. A., Krolik, J. H. 1988, *Astrophys. J.* ,**333**, 646.
Fan J. H., Xie G. Z., Pecontal E. *et al.* 1998, *Astrophys. J.*, **507**, 178.
Jurkevich I. 1971, *Astrophys. Space Sci.*, **13**, 154.
Paltani, S., Courvoisier, T. J.-L., Blecha, A., Bratschi, P. 1997, *Astron. Astrophys.*, **327**, 539.
Rutman, J. 1978, *IEEE Proceedings*, **66**, 1048.
Simonetti, J. H., Cordes, J. M., Heeschen, D. S. 1985, *Astrophys. J.*, **296**, 46.