A6523
Signal Modeling, Statistical Inference and Data Mining in Astrophysics
Spring 2015
http://www.astro.cornell.edu/~cordes/A6523

Lecture 15
– Red noise, leakage, missing information
– Ad hoc approaches
– New paradigms:
  • Entropy and information
  • Autoregressive processes and maximum entropy spectral estimation
Red Noise: Challenges for Spectral Estimation
A Bayesian test for periodic signals in red noise

S. Vaughan*

X-Ray and Observational Astronomy Group, University of Leicester, Leicester, LE1 7RH, U.K.

Accepted 2009 October 14. Received 2009 October 14; in original form 2009 September 11

ABSTRACT

Many astrophysical sources, especially compact accreting sources, show strong, random brightness fluctuations with broad power spectra in addition to periodic or quasi-periodic oscillations (QPOs) that have narrower spectra. The random nature of the dominant source of variance greatly complicates the process of searching for possible weak periodic signals. We have addressed this problem using the tools of Bayesian statistics; in particular using Markov chain Monte Carlo techniques to approximate the posterior distribution of model parameters, and posterior predictive model checking to assess model fits and search for periodogram outliers that may represent periodic signals. The methods developed are applied to two example datasets, both long XMM-Newton observations of highly variable Seyfert 1 galaxies: RE J1034 + 396 and Mrk 766. In both cases a bend (or break) in the power spectrum is evident. In the case of RE J1034 + 396 the previously reported QPO is found but with somewhat weaker statistical significance than reported in previous analyses. The difference is due partly to the improved continuum modelling, better treatment of nuisance parameters, and partly to different data selection methods.

Key words: Methods: statistical – Methods: data analysis – X-rays: general – Galaxies: Seyfert

Presents a nice summary of Bayesian methodology + comparison with periodograms.

Applies to time series with red noise + periodicity but does not deal with spectral leakage for steep spectra.
The impact of red noise in radial velocity planet searches: Only three planets orbiting GJ581?

Roman V. Baluev*

Central (Pulkovo) Astronomical Observatory of Russian Academy of Sciences, Pulkovskoe sh. 65/1, St Petersburg 196140, Russia
Sobolev Astronomical Institute, St Petersburg State University, Universitetskiy prospekt 28, Petrodvorets, St Petersburg 198504, Russia

Received 2012 November 21; in original form 2012 September 14

ABSTRACT

We perform a detailed analysis of the latest HARPS and Keck radial velocity data for the planet-hosting red dwarf GJ581, which attracted a lot of attention in recent time. We show that these data contain important correlated noise component (“red noise”) with the correlation timescale of the order of 10 days. This red noise imposes a lot of misleading effects while we work in the traditional white-noise model. To eliminate these misleading effects, we propose a maximum-likelihood algorithm equipped by an extended model of the noise structure. We treat the red noise as a Gaussian random process with exponentially decaying correlation function.

Using this method we prove that: (i) planets b and c do exist in this system, since they can be independently detected in the HARPS and Keck data, and regardless of the assumed noise models; (ii) planet e can also be confirmed independently by the both datasets, although to reveal it in the Keck data it is mandatory to take the red noise into account; (iii) the recently announced putative planets f and g are likely just illusions of the red noise; (iv) the reality of the planet candidate GJ581 d is questionable, because it cannot be detected from the Keck data, and its statistical significance in the HARPS data (as well as in the combined dataset) drops to a marginal level of \( \sim 2\sigma \), when the red noise is taken into account.

Therefore, the current data for GJ581 really support existence of no more than four (or maybe even only three) orbiting exoplanets. The planet candidate GJ581 d requests serious observational verification.

Key words: planetary systems - stars: individual: GJ581 - techniques: radial velocities - methods: data analysis - methods: statistical - surveys

Detection of periodic signals in radial velocity time series (exoplanet signatures)

His use of ‘red noise’ is a bit of a misnomer: it is correlated noise but not noise with a power-law spectrum \( \sim f^n \).
Simulating power-law noise

- Applications: many phenomena in nature are processes with spectra that are power-law in form (temporally or spatially)
  \[ S(f) \propto f^{-\alpha} \quad f_0 \leq f \leq f_1 \]

- Can use a linear filter with impulse response \( h(t) \):

  \[
  x(t) \rightarrow h(t) \rightarrow y(t)
  \]

  \[
  y(t) = x(t) \ast h(t)
  \]

  \[
  \tilde{Y}(f) = \tilde{X}(f) \times \tilde{H}(f)
  \]

  \[
  \langle|\tilde{Y}(f)|^2\rangle = \langle|\tilde{X}(f)|^2\rangle|\tilde{H}(f)|^2
  \]

- Let \( x(t) \) = realization of white noise and \( H(f) = \text{sqrt(shape)} \) of spectrum that is wanted
Procedure

• Generate spectrum in frequency domain
• Generate white noise realizations for real and imaginary parts of $H(f)$
  – random number generator: gaussian, uniform
• Multiply by $f^{-\alpha/2}$ for $f_0 \leq f \leq f_1$
• Fill vector to force Hermiticity
• Do inverse DFT to get real time-domain realization of the noise
• Can be applied to any shape of spectrum of course
• What will the statistics be in the time domain?
$f^3$ spectrum
Spectral index = 3.0  
uniform sampling

Data

Spectrum = |FFT|^2

k
Spectral index = 3.0
gaps

Data

n
Spectral index = 3.0 gaps + zerofill

Data

Spectrum = |FFT|^2
Spectral index = 3.0 before end matching

Data

\begin{align*}
\text{Spectrum} &= |\text{FFT}|^2
\end{align*}
Spectral index = 3.0 with slope lines

Data

Spectrum = $|\text{FFT}|^2$

Slope = -3
Slope = -2
Underlying Issue
The time series is periodic $\rightarrow$ discontinuities

Discontinuities
$\downarrow$
Gibbs phenomenon
$\downarrow$
Leakage
$\downarrow$
Bias in the spectral estimate
Ad hoc fix = “End Matching”

Spectral index = 3.0 with slope lines

Subtract line that connects the end points. This makes the periodically extended time series continuous.
Spectral index = 3.0 end-match line

Data

Spectrum = |FFT|^\text{2}
Spectral index $= 3.0$  after end matching
$f^5$ spectrum
Spectral index = 5.0
Spectral index = 5.0  gaps + zerofill

Data

Spectrum = |FFT|^2
Spectral index = 5.0 before end matching

Data

Spectrum = |FFT|^2

n

k
Spectral index = 5.0 with slope lines

Data

Spectrum = |FFT|^2
Spectral index = 5.0  end-match line

Data

\[ n \]

\[ \text{FFT} \]

Spectrum = |FFT|^2

\[ k \]
Spectral index = 5.0 after end matching

Data

\begin{align*}
&\text{Slope} = -5 \\
&\text{Slope} = -2
\end{align*}

\rightarrow \text{End matching not working so well}
$f^6$ spectrum
Spectral index = 6.0 uniform sampling
Spectral index = 6.0
gaps

Data

n

0 200 400 600 800 1000

0.06

0.04

0.02

0.00

-0.02
Spectral index = 6.0  gaps + zerofill

Data

Spectrum = |FFT|^2
Data Spectral index = 6.0 before end matching

Spectrum = $|\text{FFT}|^2$

n

k

$10^{-1}$ $10^{-2}$ $10^{-3}$ $10^{-4}$ $10^{-5}$ $10^{-6}$
Spectral index = 6.0 with slope lines

Data

Spectrum = |FFT|^2

n

k
Spectral index = 6.0  end-match line

Data

\[ n - 0.08 \]
\[ 0.06 - 0.02 \]
\[ 0.04 - 0.00 \]
\[ -0.02 - 0.00 \]
\[ -0.04 - 0.02 \]
\[ -0.06 - 0.04 \]
\[ -0.08 - 0.06 \]

Spectrum = \[ |FFT|^2 \]

\[ 10^{-1} \]
\[ 10^{-2} \]
\[ 10^{-3} \]
\[ 10^{-4} \]
\[ 10^{-5} \]
\[ 10^{-6} \]

\[ k \]

\[ 10^0 \]
\[ 10^1 \]
\[ 10^2 \]

Slope = -2
Slope = -6
Spectral index = 6.0 after end matching

End matching is failing!
Why is end-matching failing?

• Clue: end-matching is progressively worse for steeper spectra
Why is end-matching failing?

• Clue: end-matching is progressively worse for steeper spectra
• Answer:
   Though end matching makes the time series continuous (no discontinuities) it does not prevent discontinuities in the derivatives of the time series. As the spectrum gets steeper, the higher order derivatives become more important
Another ad hoc fix

• Tapering of the time series
• E.g. a Gaussian function centered on the middle of the time series will make the time series and its derivatives continuous (to the extent that the Gaussian has fallen off at the edges of \([0, T]\))

(Sketch on board)
3. Computer Experiments

Using the program of McClellan et al. (1979), we designed filters whose squared magnitude had a power law response in the frequency range 0.01 to 0.5, with indices of 0.5 to 5 in steps of 0.5 (10 filters in all). The impulse response of the filter with index 2.0 is shown in Fig. 1(a). The impulse response is symmetric about the $t = 0$ point, and thus the frequency response is zero phase; that is, it is purely real. The frequency response is the Fourier transform of the impulse response, and it may be approximated as closely as desired by using an FFT of the impulse response augmented by a sufficiently large number of zeroes. The process of augmenting a discrete function with a large number of zeroes before performing a finite Fourier transform is called zero padding and results in a closer spacing of the transformed

$n = 2$

Figure 1. (a) Impulse response of a filter designed to have a squared power law frequency response with a slope of -2. A total of 201 weights are used. (b) Squared frequency response of the filter. The slope is -2.0000. (c) The 1000-point sample of Gaussian white noise used as input to the filter. (d) The 800-point output; at each end, 100 points are lost because the symmetrical filter is 201 weights long.
Figure 2. (a) Maximum entropy spectrum of the signal from Fig. 1(d). The ordinate is $10 \log_{10} (\text{PSD})$. (b) Periodogram of the same signal.

When the spectral index is increased to 4, the resulting impulse response, squared frequency response, white noise input, and red noise realization are those given in Fig. 3. The MEM spectrum, this time using 10 weights, and the periodogram are given in Fig. 4.
Figure 3. (a) Impulse response of a filter designed to have a squared power
law frequency response with a slope of $-4$. A total of 301 weights are used.
(b) Frequency response of the filter. The slope is $-4.0009$. (c) The 1000-
point sample of Gaussian white noise used as input to the filter. (d) The
700-point output. At each end, 150 points are lost because the symmetrical
filter is 301 weights long.

Figure 4. (a) Maximum entropy spectrum of the signal from Fig. 3(d). The
ordinate is $10 \log_{10} (\text{PSD})$. (b) Periodogram of the same signal.
Figure 5. Observed FFT (periodogram) index versus the MEM index (index is the negative of the slope). There are 100 independent realizations of the power law process for each of the 10 indices 0.5, 1.0, ..., 5.0. In every case the same time series was used as input to MEM and to FFT (periodogram). (a) Raw data. (b) End-matched data. (c) Windowed data. (d) End-matched and windowed data.
Cyclical Data

Other workarounds

• Convolution of a short filter function with a long time series (too long to use a single FFT)

• Cross correlation of two long time series
A guided tour of the fast Fourier transform

The fast Fourier transform algorithm can reduce the time involved in finding a discrete Fourier transform from several minutes to less than a second, and also can lower the cost from several dollars to several cents

G. D. Bergland  Bell Telephone Laboratories, Inc.

For some time the Fourier transform has served as a bridge between the time domain and the frequency domain. It is now possible to go back and forth between waveform and spectrum with enough speed and economy to create a whole new range of applications for this classic mathematical device. This article is intended as a primer on the fast Fourier transform, which has revolutionized the digital processing of waveforms. The reader’s attention is especially directed to the IEEE Transactions on Audio and Electroacoustics for June 1969, a special issue devoted to the fast Fourier transform.

dcrete version of the Fourier transform (DFT) that must be understood and used. Although most of the properties of the continuous Fourier transform (CFT) are retained, several differences result from the constraint that the DFT must operate on sampled waveforms defined over finite intervals.

The fast Fourier transform (FFT) is simply an efficient method for computing the DFT. The FFT can be used in place of the continuous Fourier transform only to the extent that the DFT could before, but with a substantial reduction in computer time. Since most of the problems associated with the use of the fast Fourier transform actually stem from an incomplete or incorrect under-
Cyclical vs Non-cyclical Convolution & Correlation

From Bergland, “A Guided Tour of the FFT”, IEEE Spectrum, 1969, 6 (Bell Labs)
Noncyclical Procedures

**Figure 16.** Noncyclical convolution of two finite signals analogous to that performed by the FFT algorithm.

**Figure 17.** A method for convolving a finite impulse response with an infinite time function by performing a series of fast Fourier transforms.

**Figure 18.** A method of using the fast Fourier transform algorithm to compute $N$ lags of the autocorrelation function of an $M$-term series.
Aspects of Spectral Analysis

The outcome of any attempt to estimate the power spectrum should be assessed through consideration of the following:

1. confidence intervals
2. resolution
3. handling of gaps (missing data)
4. bias (particularly for processes with steep power laws)

Numbers (2) and (3) are related because resolution is determined by the length of the data set $T$ and one can view data after the interval $[0, T]$ as missing data.
Linear/fixed window FT methods vs. Data adaptive methods

Spectral estimators based on the Fourier transform have the properties:

no smoothing

\[ \begin{align*}
1. & \text{ frequency resolution: } \delta f = T^{-1} \quad T = \text{data span} \\
2. & \text{ estimation error: } \varepsilon = \frac{\text{Var} \hat{S}}{\langle \hat{S}(\omega) \rangle} = 1
\end{align*} \]

with smoothing

\[ \begin{align*}
3. & \text{ tradeoff between } \varepsilon \text{ and } \delta f: \\
   & \text{ spectral window } \tilde{W} \Rightarrow \varepsilon \text{ decreases, } \delta f \text{ increases,} \\
   & \text{ sidelobes altered } \varepsilon = \left( \frac{2}{N_{\text{d.o.f.}}} \right)^{1/2}
\end{align*} \]

4. Since \( \hat{R}_w(\tau) \Leftrightarrow \hat{S}_w(\omega) \), the estimator is linear in the correlation function estimates, \( \hat{R}_w \).

5. Also, once chosen by the analyst, the window function is fixed with respect to the data.

Therefore, even though the spectral estimation may be iterative whereby the analyst chooses by trial an optimum window, this optimization is performed according to intuitive/aesthetic/prior knowledge prejudice. The form of the window is not an integral part of the estimation process.

Hence, the characterization of finite F.T. estimators as linear, fixed window, empirical estimators.
Different Views on Missing Data in a Time Series

1. Transform (in the general sense) data using fixed window analyses; then correct the data for the effects of gaps in data, etc. (e.g. can apply a CLEAN algorithm to the spectrum).

   vs.

2. Transform the data without making assumptions about missing data (maximum entropy techniques are “maximally noncommittal” about missing data).

‘Empirical methods such as Blackman-Tukey, which do not invoke even a likelihood function, are useful in the preliminary, exploratory phase of a problem where our knowledge is sufficient to permit intuitive judgments about how to organize a calculation (smoothing, decimation, windows, prewhitening, padding with zeroes, etc.) but insufficient to set up a quantitative model which would do the proper things for us automatically and optimally.’ [In abstract of ”On the Rationale of Maximum Entropy Methods”, E.T. Jaynes, 1982, IEEE Proc., 70, 939.]