Data Analysis: Time Series

Steve Schwartz

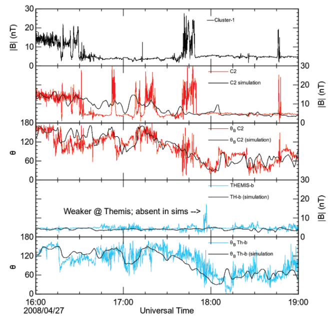
Imperial College London

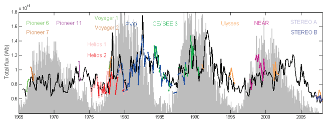
4 December 2017

Contents

Basics - Time Series

- In situ space data, sunspot numbers, mean temperatures, etc. are all Time Series $\Rightarrow t$ is independent variable

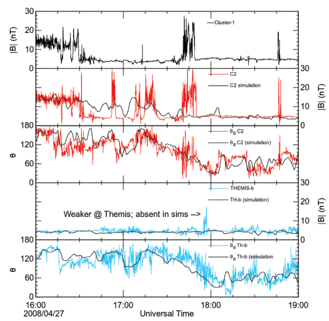




See *Analysis Methods for Multi-spacecraft Data*, ISSI Scientific Report, SR-001, ISSI, Bern, Switzerland, 1998 and more recent ... *Revisited*

Basics - Time Series

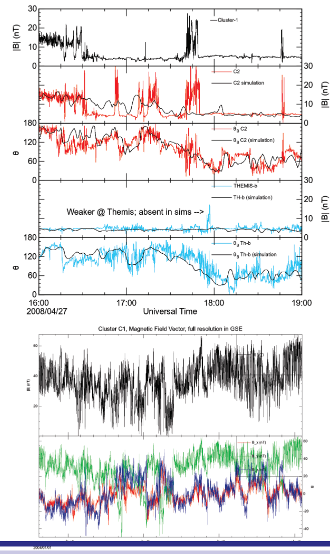
- In situ space data, sunspot numbers, mean temperatures, etc. are all Time Series $\Rightarrow t$ is independent variable
- Simple plots often convey the information you want



See *Analysis Methods for Multi-spacecraft Data*, ISSI Scientific Report, SR-001, ISSI, Bern, Switzerland, 1998 and more recent ... *Revisited*

Basics - Time Series

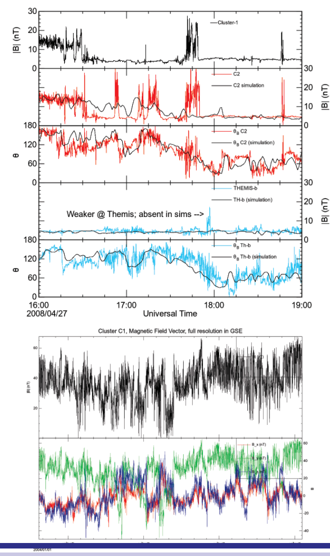
- In situ space data, sunspot numbers, mean temperatures, etc. are all Time Series $\Rightarrow t$ is independent variable
- Simple plots often convey the information you want
- But not always - science you do depends on resolution of data you use



See *Analysis Methods for Multi-spacecraft Data*, ISSI Scientific Report, SR-001, ISSI, Bern, Switzerland, 1998 and more recent ... *Revisited*

Basics - Time Series

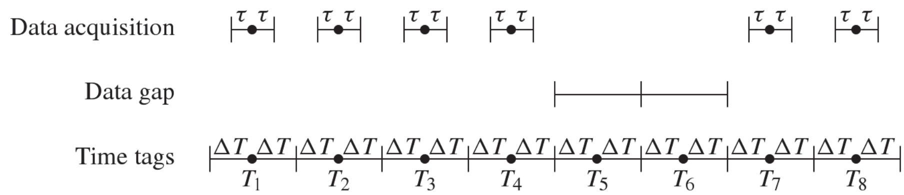
- In situ space data, sunspot numbers, mean temperatures, etc. are all Time Series $\Rightarrow t$ is independent variable
- Simple plots often convey the information you want
- But not always - science you do depends on resolution of data you use
- Extracting information involves a variety of manipulations and specialised techniques



See *Analysis Methods for Multi-spacecraft Data*, ISSI Scientific Report, SR-001, ISSI, Bern, Switzerland, 1998 and more recent ... *Revisited*

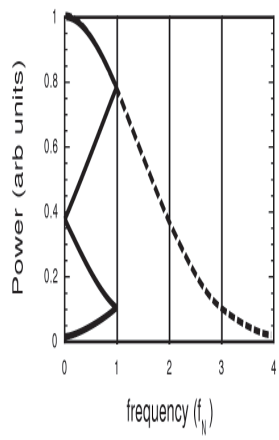
Properties - Time Sampling

- Data sampled at discrete Sample Intervals $2\Delta T$ or Sample Frequency $\equiv f_s = 1/(2\Delta T)$
- Each datum valid for/average over/... Sample duration $= 2\tau$
- Real time series have Data Gaps where you would otherwise expect a measurement
- Note: not all time series have Regular (equally-spaced) timetags



Nyquist Theorem and Aliasing

- Often useful to pass from time to frequency domain
- Limiting frequency is Nyquist frequency $\equiv F_N = f_s/2 = 1/(4\Delta T)$
- Worse still, consider signals $\cos 2\pi(f_N - f)t$ and $\cos 2\pi(f_N + f)t$
- If sample these at f_s (i.e., at $t_n = 2n\Delta T$) they give same result!
- Hence power above f_N is folded into range $0 < f < f_N$
- Two lessons:
  - Lose factor 2 from f_s to f_N
  - Filter prior to sampling to limit aliasing



Resampling of Time Series

Resampling or "joining" means putting a time series (or a continuous analogue signal) onto a new set of timetags.
Why do this?

- Attenuate higher frequency components to limit aliasing
- Remove higher frequency components to reveal longer timescale trend
- Determine and subtract trend ("de-trend")
- Ensure regular timetags for, e.g., Fourier analysis
- Match two time series (e.g., monthly rainfall and daily temperature) to enable correlation analyses

Pitfalls:

- Nyquist (every touch loses factor of 2)
- Data gap handling; Data end handling
- Introduction of spurious points

Increasing resolution

When new "target" timetags are more closely-spaced than original data, need some method of interpolation to increase time resolution

Fuzzy Join or "nearest neighbour" takes the original datum closest to new timetag

Linear interpolation connects data by straight lines to find values at new timetags. This is safe, but retains sharp corners

Spline interpolation fits a curve to the data, ensuring continuous *derivatives* at the expense of generating loops and other artefacts (e.g., negative temperatures)

Gap/end handling:

- Interpolate/extrapolate anyway
- Remove (can't do if want to take FFT)
- Zero fill or pad (e.g., nearest neighbour)
- Gaussian end fill

You can NOT increase information, but it is easy to fool yourself

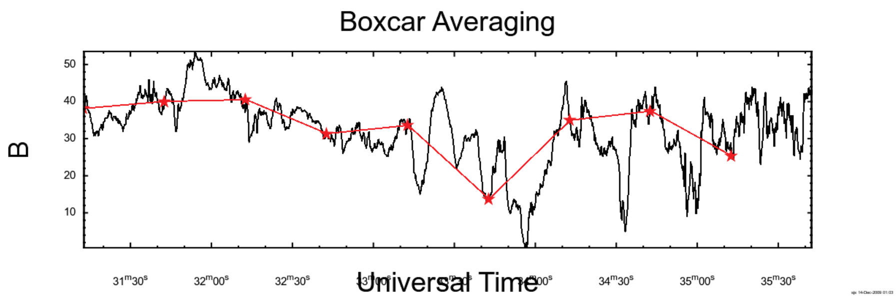
Decreasing resolution - Boxcar

When new "target" timetags are more widely-spaced than original data, need some method of averaging to decrease time resolution. Can, but usually shouldn't, use interpolation methods

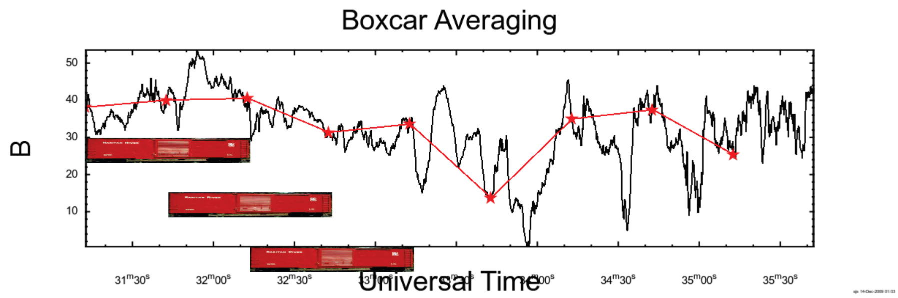
Boxcar Averaging defines a ΔT_{box} around the new timetags and performs a simple average.

- Good especially if lots of points in boxcar
- To avoid too much aliasing requires $\Delta T_{\text{box}} > 2\Delta T_{\text{target}}$ (i.e., boxcars overlap one another $> 50\%$)
- Use $\Delta T_{\text{box}} = 2\Delta T_{\text{target}}$ to retain maximum information
- Use larger ΔT_{box} for more smoothing

Boxcars in action



Boxcars in action



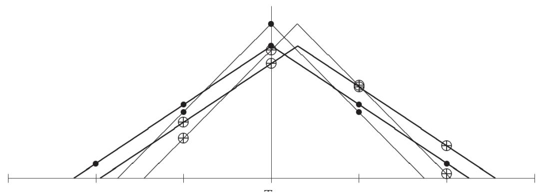
Decreasing resolution - Other averages

Bartlett window uses triangular weighting and can improve aliasing for fewer points within window

Gaussian window is not normally used as it suppresses signal *below* f_N

Gap/end handling - all averaging techniques:

- Interpolate/extrapolate
- Remove (can't do if want to take FFT)
- Zero fill or pad (e.g., nearest neighbour)
- Gaussian end fill



Different width and placed Bartlett windows

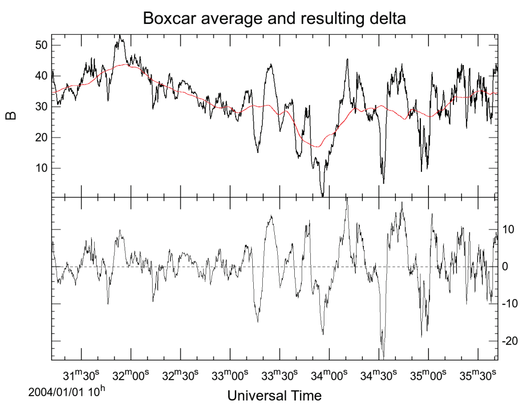
Resampling same/similar resolutions

- This is hard
- Boxcar introduces phase errors
- Linear interpolation introduces amplitude errors (beats appear)
- Bartlett may be best, but often linear interpolation is used and is adequate

Detrending and smoothing

Resampling techniques can be used to find and remove trends

Before many analyses (e.g., minimum variance) it is critical to pre-prepare the data.

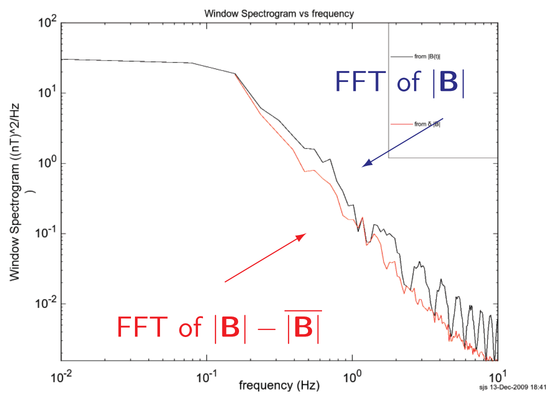


Time Series Analysis - Power Spectra

Fast Fourier Transform - FFT - most widely used method

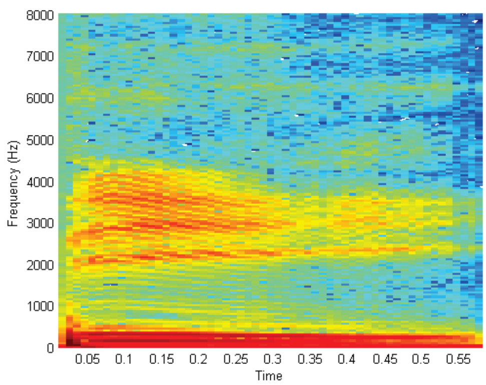
Requires 2^n data points, so requires preselection

Influenced by trends (below uses data from previous slide)



Spectrograms

By taking FFT's at regular intervals one can build up a spectrogram showing how the frequency content of the signal changes with time.



Time Series Analysis - Wavelet Transforms

Fourier techniques convolve the signal with a sine curve, e.g.,

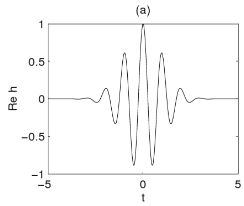
$$A(f_n) \propto \int Y(t) \sin(2\pi f_n t)\ dt$$

which is less effective for isolated structures or signals that don't remain coherent.

Wavelets are localised kernels that are used instead:

Not necessarily orthogonal nor complete

Many forms (Morlet, Daubechie - below, Haar, ...); can tune to process of interest



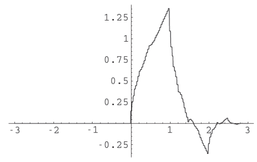
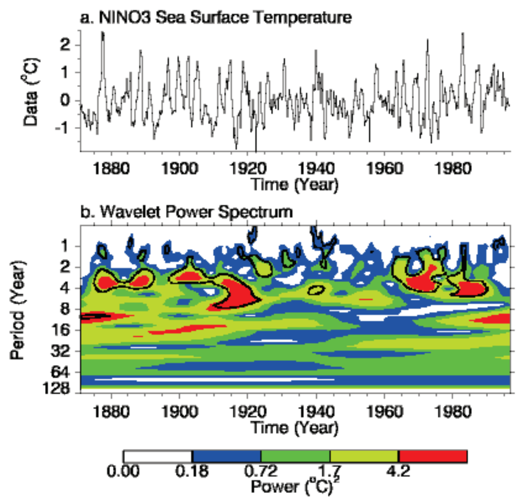


FIGURE 14. The D4 father function.

Wavelet example



Cross-Correlations

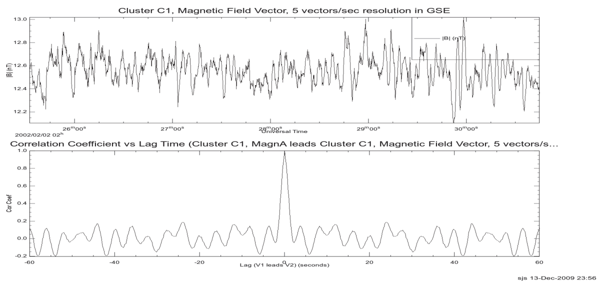
The discrete cross-correlation comes from the covariance between two variables at $r\times$ lag step:

$$C_{xy}(r) = \frac{\sum [x(i) - \overline{x}]\,[y(i+r) - \overline{y}]}{n - r}$$

normalised to $\sigma_x\sigma_y$ so that $-1 < C_{xy} < +1$

Gives information on both the frequency content and phase

C_{xx} is closely related to the power spectrum of x



Physics-based analyses

For particular physical systems there are specialised analyses, e.g.:

- Minimum variance of magnetic field ($\nabla \cdot \mathbf{B} = 0 \Rightarrow$ for planar structures one component of **B** should be constant, so find the direction in which **B** varies least)
- Walen test for Alfvén waves - various components of **B** and **V** should correlate

Summary

- Time series need to be handled with care
- Need to filter, make regular, or otherwise prepare data for further analysis
- Nyquist limits constrain information content
- Aliasing confuses frequency of signal
- Several power estimates
- Other specialised analyses
- Read Chapters 1 & 2 of ISSI book