TIME-DOMAIN ASTRONOMY

Lectures 3: Spectral Analysis

Stefano Covino
INAF / Brera Astronomical Observatory

Fourier Analysis

- The aim of Fourier analysis is to express any function as a sum of different sines and cosines, characterised by an angular frequency ω or a corresponding time period P = 2π/ω.
- Such a decomposition is known as a Fourier series.

$$f(t) = \frac{a(0)}{2} + \sum_{n=1}^{\infty} [a(n) \sin n\omega_0 t + b(n) \cos n\omega_0 t]$$

- This expression can represent practically any periodic function with period P_0, with suitable adjustment of the coefficients a(n) and b(n), which are known as the Fourier coefficients.

Jean Baptiste Joseph Fourier

Fourier Decomposition

- The conditions under which the Fourier decomposition is valid are that f(t) has only a finite number of finite discontinuities and only a finite number of extreme values within a period.
- These are known as Dirichlet conditions and the functions obeying them are called piece-wise regular.

- The Fourier coefficients a(n) and b(n) can be determined by performing the following integrations:

$$a(n) = \frac{2}{P_0}\int_0^{P_0} f(t)\sin n\omega_0 t\, dt \;\; ; \;\; b(n) = \frac{2}{P_0}\int_0^{P_0} f(t)\cos n\omega_0 t\, dt, \;\; n = 0, 1, 2\ldots$$

Fourier Decomposition

- Alternatively, the Fourier decomposition may also be expressed in an equivalent exponential form:

$$f(t) = \frac{1}{P_0} \sum_{n=-\infty}^{\infty} c(n) e^{in\omega_0 t}; \text{ where } c(n) = \int_0^{P_0} f(t) e^{-in\omega_0 t} dt$$

- In the above expansions, the n = 0 term is often called the constant or the D.C. (Direct Current) component, the n = 1 term the fundamental and the terms with n > 1 the harmonics.

Fourier Decomposition

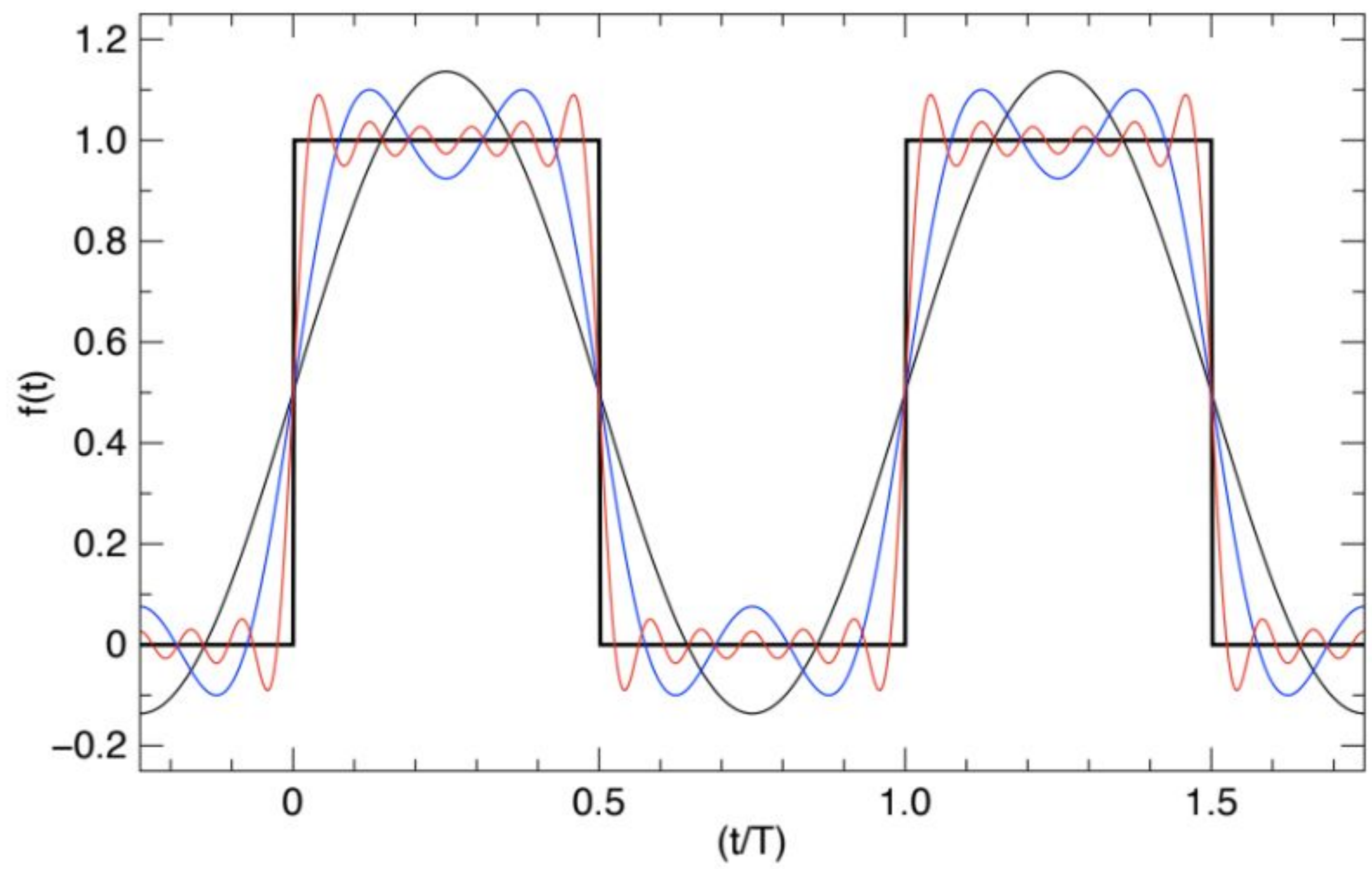


Fig. 1: Example of a Fourier series decomposition. The original time domain function $f(t)$, shown in the thick black line, is a square wave of period T, which equals unity from $t = 0$ to $0.5T$ and zero from $0.5T$ to T. Two periods of the function are plotted. Thin lines show the Fourier sum with different number of terms: (i) black: DC+Fundamental, (ii) blue: sum up to 3rd harmonic, (iii) red: sum up to 11th harmonic. The gradual improvement in the approximation of the function with increasing number of terms in Fourier series is evident. At the point of discontinuity, all the reconstructions pass through the average of the left and right limits of the original function and the transition gets progressively sharper with larger number of terms.

Continuous Fourier Transform

- We define the Fourier transform (FT) of any function f(t) as:

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt$$

- This is a linear transformation and no information is lost. The representations of a function in time and frequency domains are equivalent.

- The original f(t) can be recovered by applying the inverse Fourier transform:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) e^{i\omega t} d\omega$$

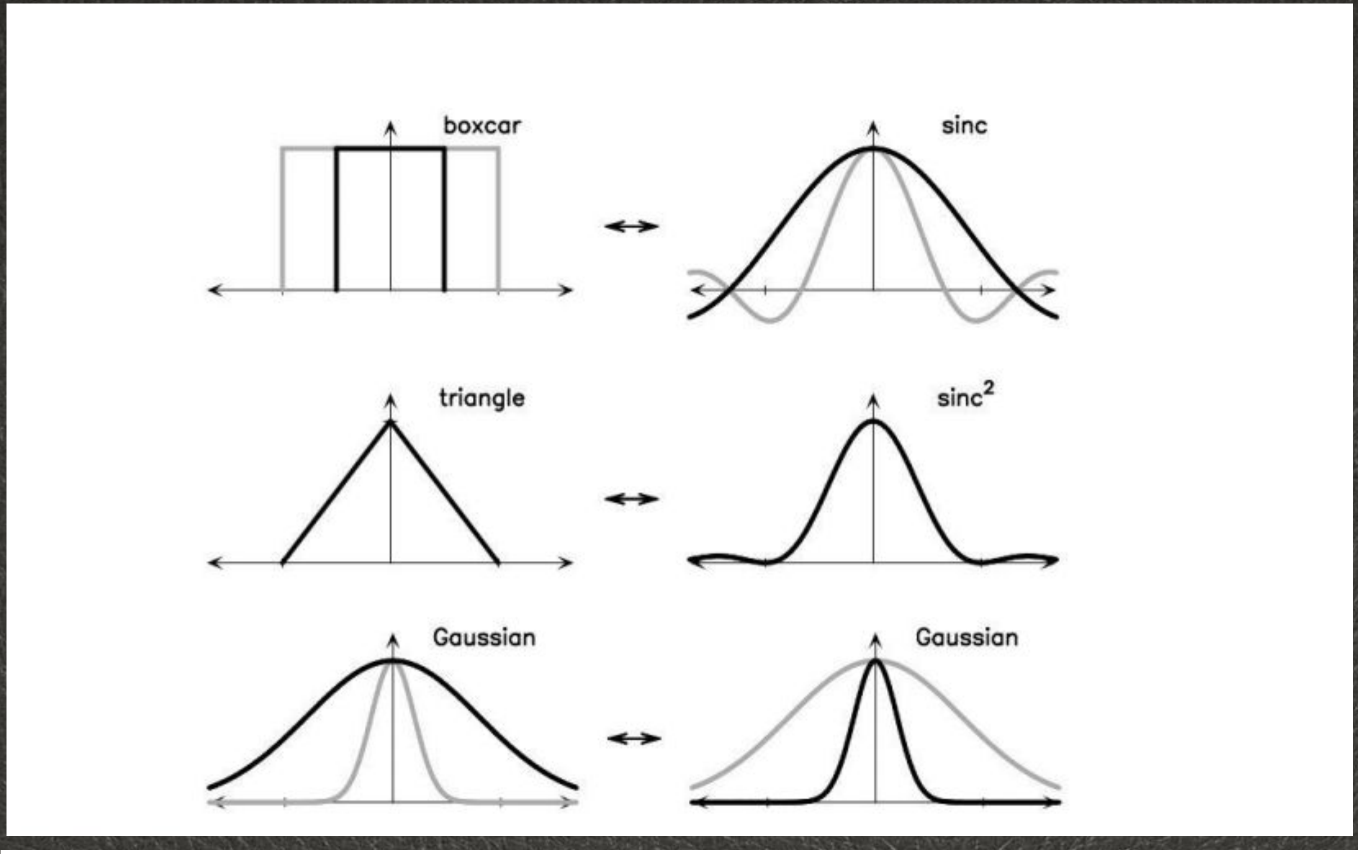
Continuous Fourier Transform

- The FT has a number of interesting properties. It is linear, not necessarily a real function, and its amplitude is invariant to time shift (but not its phase).

f(t)	F(ω))
Real	$H(-\omega) = H^*(\omega)$
Even	Even
Odd	Odd
Real & Even	Real & Even
Real & Odd	Imaginary & Odd

- Unless the original function is even, an unlikely situation in the case of a time series, its Fourier transform is complex!

Fourier Transform Examples



Scaling; "broad ← → narrow"

h(at) ← → H(f/a) / |a|

Shifting; "shift ← → phase roll/gradient"

$h(t-t_0)$ ← → $H(f) * \exp(2\pi i\, f\, t_0)$

Convolution; "convolution ← → multiplication"

h(t) * g(t) ← → H(f) G(f)

Power Spectral Density

- The Power Spectral Density (PSD) is defined as the Fourier Transform multiplied by its complex conjugate and therefore the square modulus of the Fourier Transform:

$$P(\omega) = F(\omega) \cdot F^*(\omega) = |F(\omega)|^2$$

- If the original function is real, which is of course usually the case for time series, the PSD is an even function and the values at negative frequencies are redundant.

Power Spectral Density

- The FT is a linear function, the PSD is not. This means that while the FT of the sum of two signals is the sum of the FT of the signals, in the case of the PSD this is not true and there are cross-terms to be considered.

- E.g., if two signals are f(t) and g(t), the PSD of the sum of the two is:

$$P[f(t)+g(t)] = |F[f(t)+g(t)]|^2 = F[f(t)+g(t)] \cdot F^*[f(t)+g(t)] =$$
$$= P[f(t)] + P[g(t)] + 2Re\{F[f(t)] \cdot F[g(t)]\}$$

- If the two signals are uncorrelated, the cross-term is zero and linearity applies.

Autocorrelation function

- The autocorrelation (ACF) of a function f(t) (more later through the course) is defined as:

$$A(t) = \int_{-\infty}^{+\infty} f(\tau) f(t+\tau) d\tau \Longleftrightarrow F(f)F^*(f) \equiv |F(f)|^2$$

- The autocorrelation of a function is the Fourier transform of its PSD.

- It is also simple to derive Parseval's theorem (simply setting t=0):

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |f(\omega)|^2 d\omega$$

PSD and ACF are Fourier pairs

- Useful notebook:

1. PSD&ACF

Discrete Fourier Transform

- In the real world we have discrete measurements extending from 0 to T: a time series (commonly called, in astronomy, "light curve") consisting of N measurements x_k taken at equally-spaced times t_k from 0 to T.
- In this case we can define the discrete Fourier transform (and its inverse) as:

$$a_j = \sum_{k=0}^{N-1} x_k e^{-2\pi i jk/N} \qquad (j = -N/2, \ldots, N/2-1)$$

$$x_k = \frac{1}{N} \sum_{j=-N/2}^{N/2-1} a_j e^{2\pi i jk/N} \qquad (k = 0, \ldots, N-1)$$

Discrete Fourier Transform

- Since the data are equally spaced, the times are kT/N and the frequencies are j/T.
- The time step is $\delta t = T/N$ and the frequency step is $\delta \nu = 1/T$.
- As the discrete time series has a time step δt and a duration T, there are limitations to the frequencies that can be examined.
  - The lowest frequency is 1/T, corresponding to a sinusoid with a period equal to the signal duration.
  - The highest frequency that can be sampled, is called Nyquist frequency:

$$\nu_{Nyq} = \frac{1}{2\delta t} = \frac{1}{2}\frac{N}{T}.$$

Discrete Fourier Transform

- There is a zero frequency, at which the FT value is simply the sum of the signal values:

$$a_0 = \sum_{k=0}^{N-1} x_k e^{-2\pi i 0k/N} = \sum_{k=0}^{N-1} x_k$$

$$P_j = |a_j|^2$$

- Parseval's theorem applies also to the discrete case and one can see that the variance of the signal is 1/N times the sum of the a_j over all indices besides zero (also known as Plancherel Theorem):

$$\sum_{k=0}^{N-1} |x_k|^2 = \frac{1}{N} \sum_{j=-N/2}^{N/2-1} |a_j|^2$$

$$\mathrm{Var}(x_k) \equiv \sum_k (x_k - \overline{x})^2 = \sum_k x_k^2 - \frac{1}{N}\left(\sum_k x_k\right)^2 = \frac{1}{N}\sum_j |a_j|^2 - \frac{1}{N} a_0^2,$$

$$\mathrm{Var}(x_k) = \frac{1}{N} \sum_{\substack{j=-N/2 \\ j\neq 0}}^{N/2-1} |a_j|^2.$$

Power Spectral Density

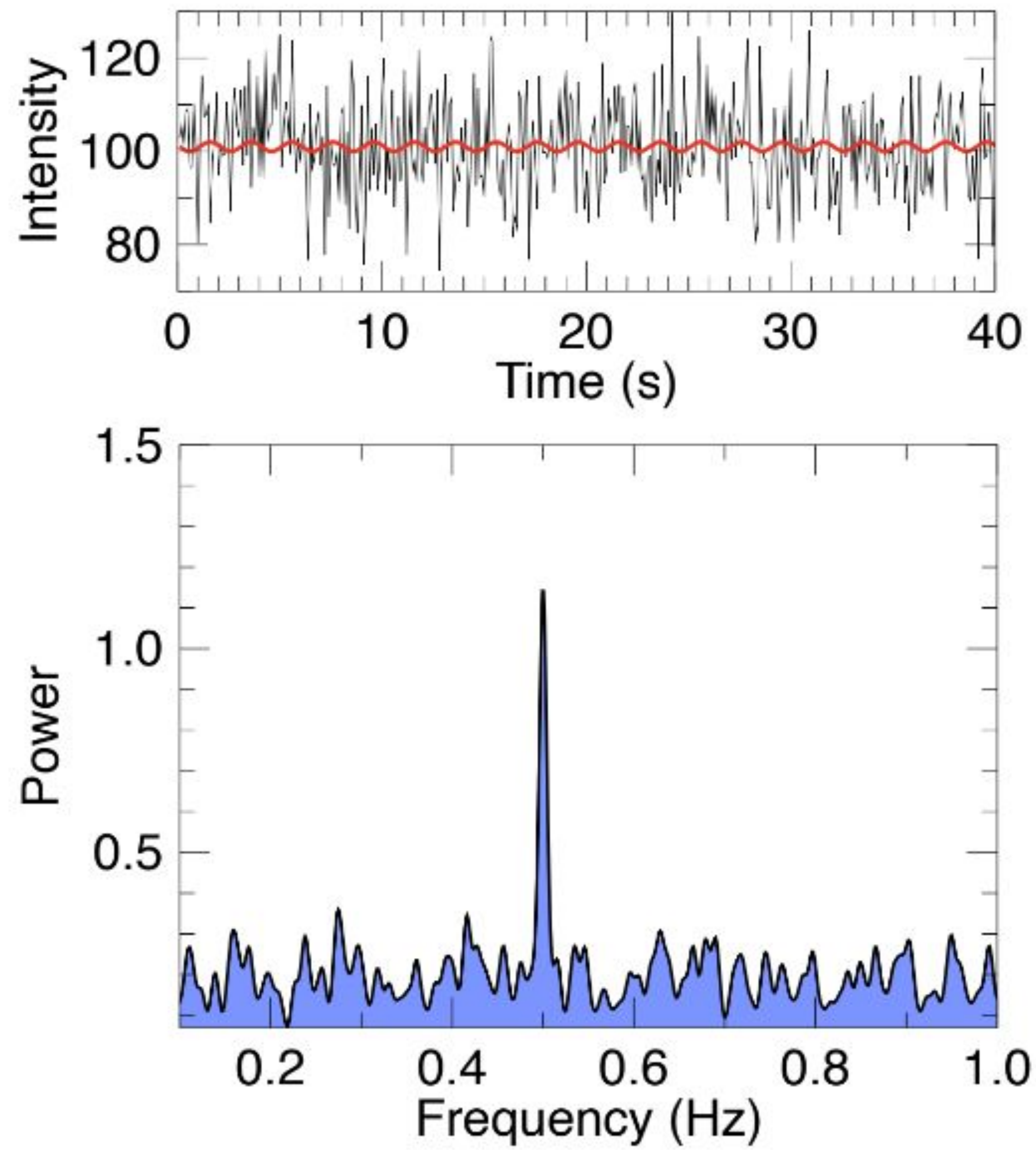


Fig. 2: Top panel: the black line shows 40 seconds of a simulated 1000-s time series consisting of a weak sinusoidal modulation (red line) "drowned" into a strong Gaussian noise. Bottom panel: corresponding PDS, zoomed on the relevant frequency range, where the modulation is clearly visible. A weak signal spread in time is collected into a single frequency bin at high significance.

Windowing and Sampling

- The CFT and the DFT can be connected easily taking into account that the FT of the product fo two functions is the convolution of the FT of the functions.

$$F[x \cdot y] = F[x] \otimes F[y] = \int_{-\infty}^{+\infty} F[x(\nu')] F[y(\nu - \nu')] d\nu'$$

- A discrete time series $x(t_k) \equiv x_k$ can be seen as the product of a continuous function f(t) over $(-\infty,\infty)$ and two additional functions: w(t) to limit it to the (0,T) interval and s(t) to sample it at times t_k:

$$x_k = h(t) \cdot w(t) \cdot s(t)$$

Windowing and Sampling

- w(t) is a boxcar window function, which is 1 in the (0,T) interval and zero outside. s(t) is a series of delta functions at t_k, spaced by T/N:

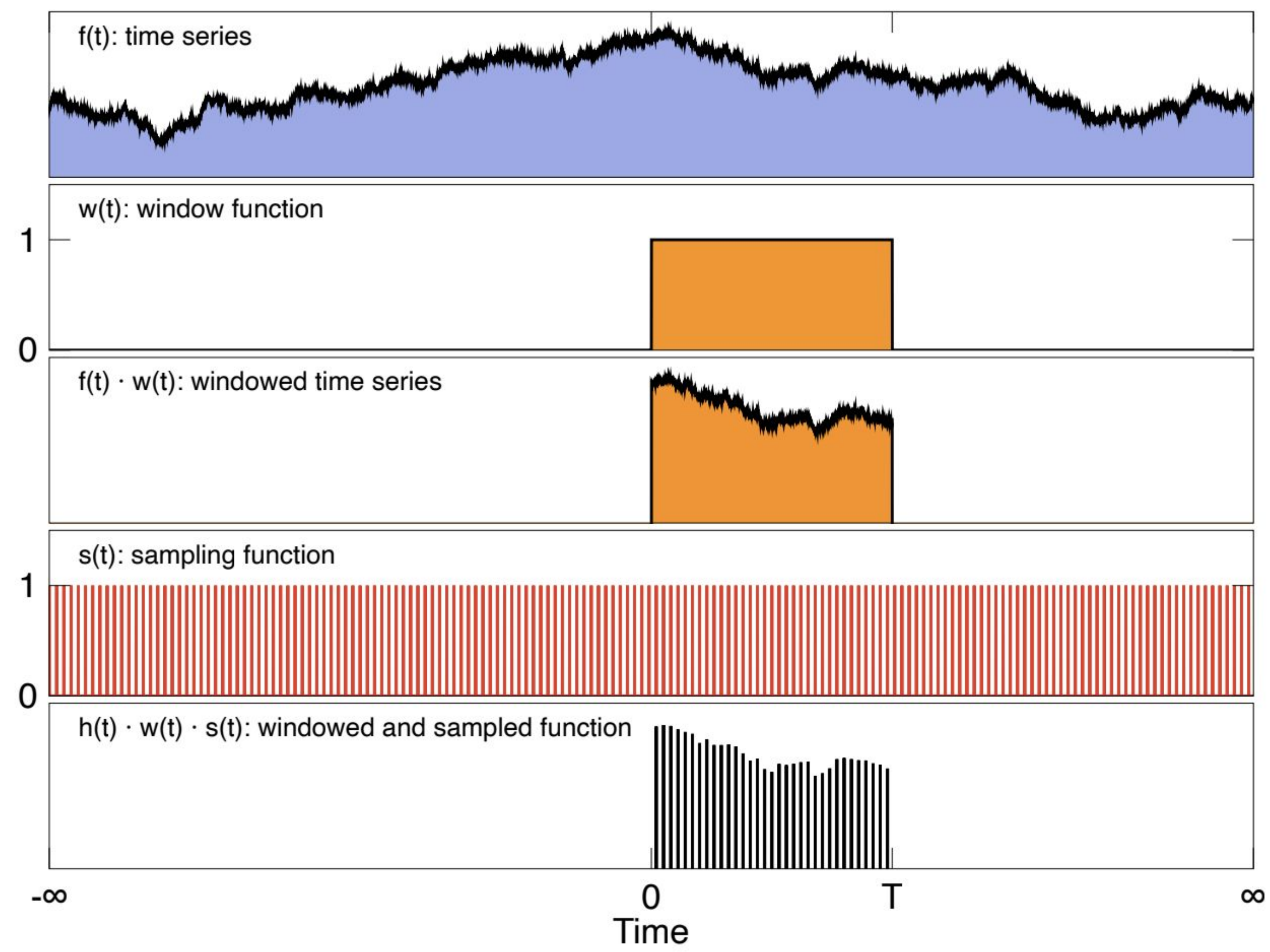


Fig. 3: From continuous function to discrete data. The panels from top to bottom : continuous function $f(t)$ over the $(-\infty,\infty)$ interval; boxcar window function $w(t)$; $f(t)\cdot w(t)$ windowed continuous function; $s(t)$ sampling function; final windowed and sampled function, corresponding to the discrete data.

Windowing effects

- Let us consider a purely sinusoidal function f(t) = sin(ωt), whose FT is a delta function at ω.
- The multiplication by the window function corresponds to the convolution of the delta function with the FT of the window.
- It is simple to calculate the FT of the window: we consider a window function that is unity in the −T/2,T/2 interval, as it is a real and even function, whose FT is also real and even:

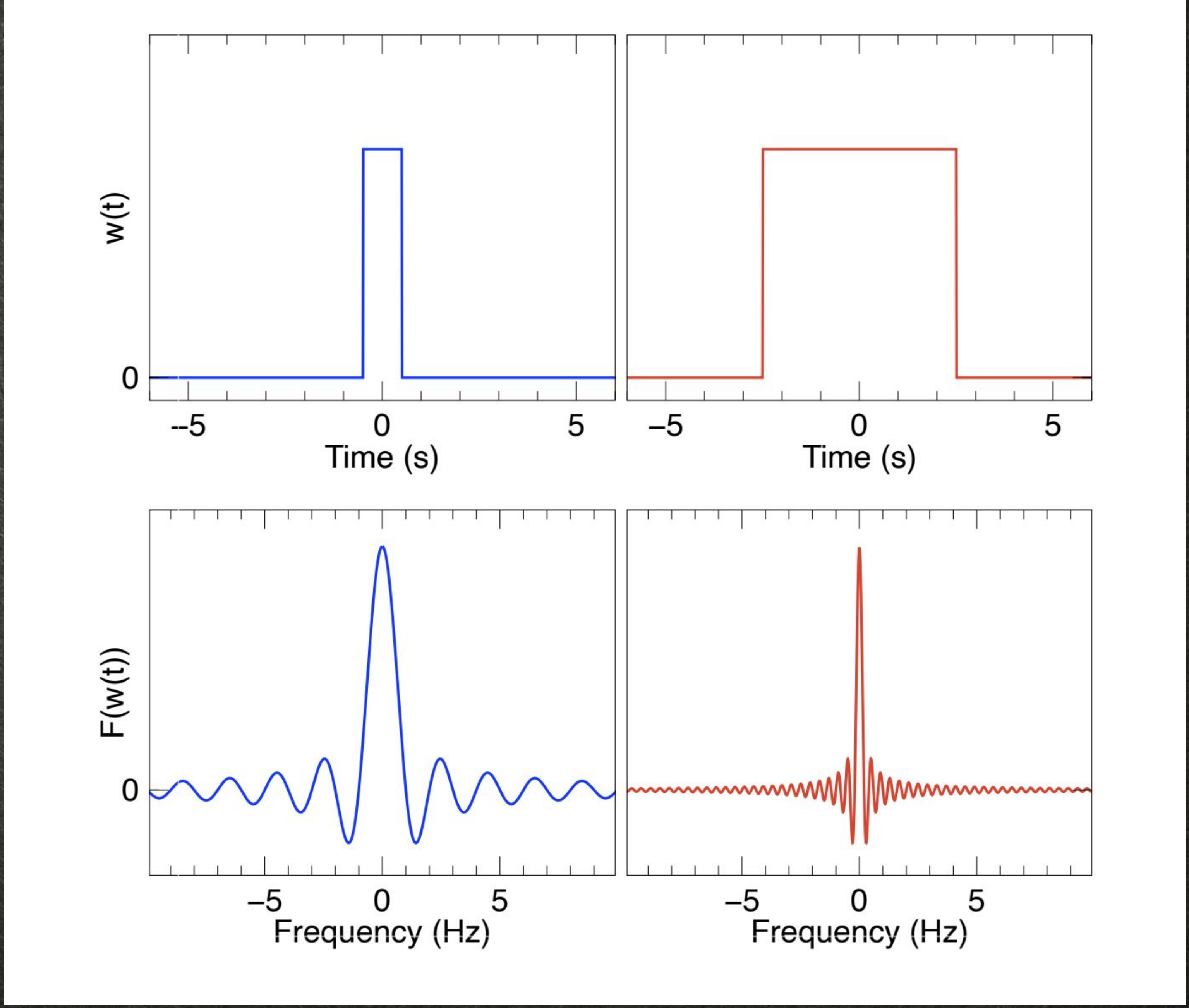
$$F(w(t)) = 2\frac{\sin(\pi \nu T)}{\pi \nu}$$

  - This is the well known "*sinc*" function.

Windowing effects

- An important general rule, the FT peak is broader for shorter T.

- The resolution of the signal FT is therefore higher the longer the observation is.
- In addition to the broadening, there is the formation of side lobes. They are much lower than the central peak, but cannot always be ignored.



Sampling effects: aliasing

- The FT of a series of regularly spaced delta functions with spacing T/N is itself a series of delta functions with spacing N/T:

$$s(t) = \sum_{k=-\infty}^{+\infty} \delta(t - \frac{kT}{N}) \Longleftrightarrow F(s(t)) = \sum_{m=-\infty}^{+\infty} \delta(\nu - \frac{mN}{T})$$

- Therefore, the effect of sampling on the FT of a sinusoidal signal with frequency ν_0 (a delta function at ν_0) is that of adding an infinite sequence of delta functions spaced by N/T, called *aliases*.

- Depending on the frequency of the original signal and the Nyquist frequency we can have different situations.

Aliasing

- This causes that features at $\nu = \nu_{N/2} + \nu_x$ also appear at $\nu = \nu_{N/2} - \nu_x$.

- Summing up:
  - the transition from the CFT to the DFT involves two operations: windowing, a convolution with the function W(f), which is essentially a peak with a width $\delta f = 1/T$ plus sidelines, and aliasing, a reflection of features above the Nyquist frequency back into the range $(0,\nu_{N/2})$.
  - Windowing is caused by the finite extent, aliasing by the discrete sampling of the data.

Sampling effects: aliasing

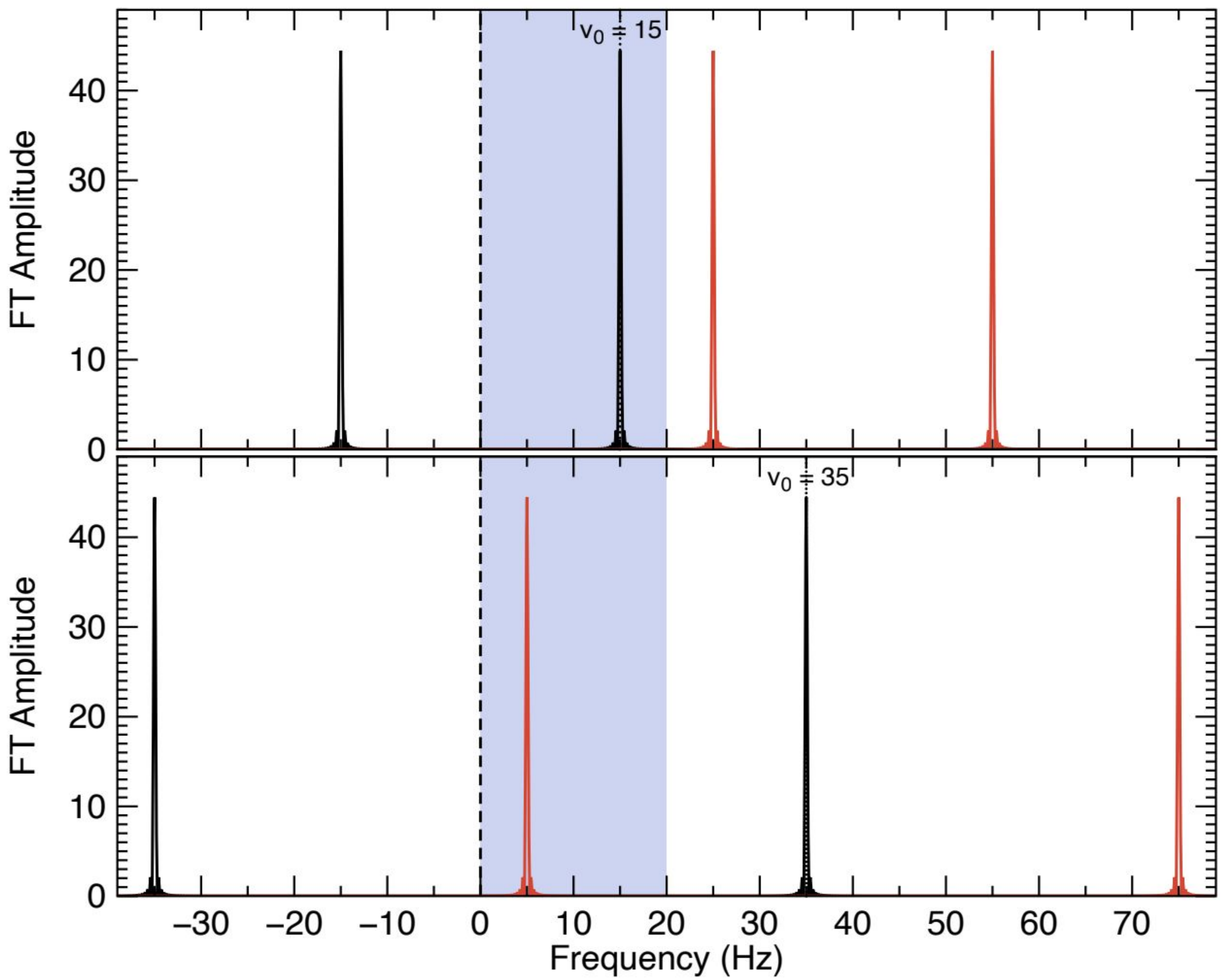


Fig. 5: Top panel: Aliasing on a sinusoidal signal at 15 Hz sampled at 40 Hz. In black the true signal FT amplitude, in red the aliased one. The blue region marks the interval below ν_{Nyq}. The signal is detected and the aliases are not. Bottom panel: same as the top panel, but with a signal at 35 Hz. Here the signal is not detected, but the 5 Hz alias is.

Sampling effects: aliasing

- We have all experienced aliasing effects when looking at fast rotating objects like an air fan under fluorescent light.
  - The light provides a sampling at 50 Hz (or 60 Hz, depending on where you live), while the fan has a periodicity.
  - Depending on its angular speed, you can see it rotating apparently much slower, or even to stop and rotate in the opposite direction.

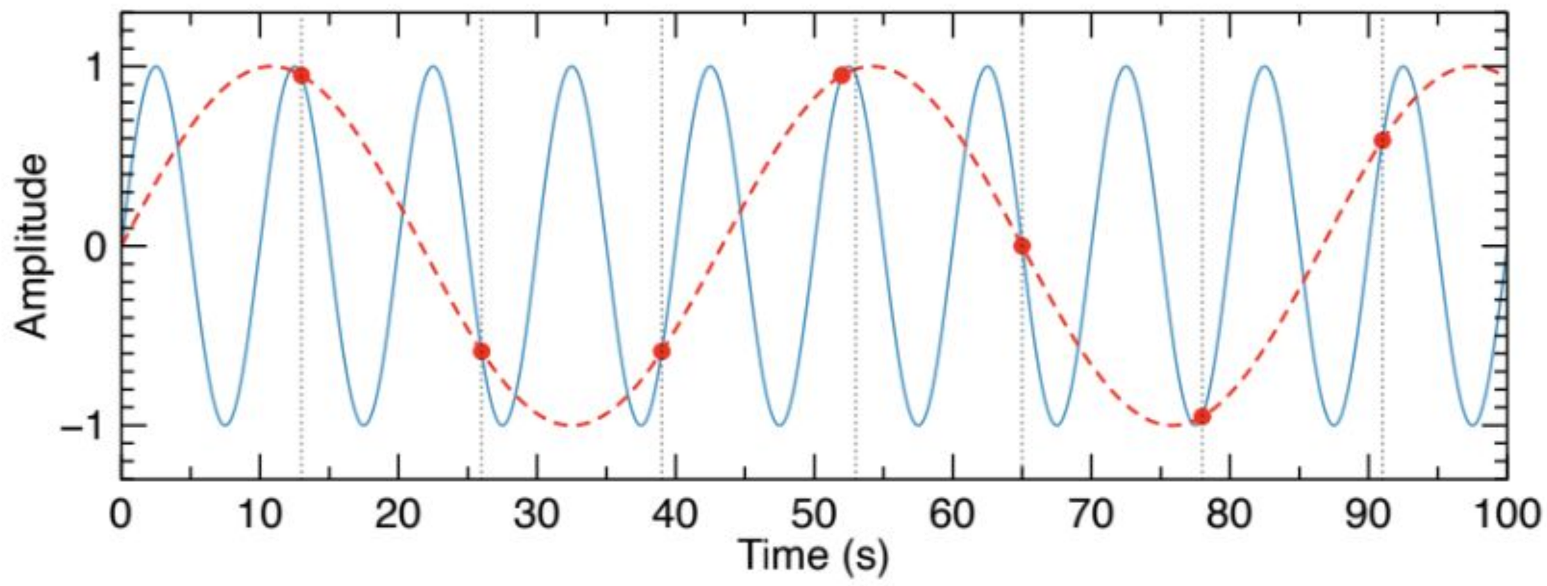


Fig. 6: Time domain example of aliasing. The blue signal has $\nu_0 = 0.1$ Hz. If it is sampled with $\nu_{Nyq} = 0.038$ Hz (red points) the red dashed alias is the best fit to the data, with $\nu_a = 0.023$ Hz.

Sampling effects: aliasing

- In the real world we do not really sample signals, but integrate them over finite time bins, i.e. we *convolve* it with a binning function:

$$b(t) = \begin{cases} \frac{N}{T} & t \in [-\frac{T}{2N}, \frac{T}{2N}] \\ 0 & \text{outside} \end{cases}$$

- Therefore, the signal FT will be multiplied by that of the binning function, which is again a *sinc* function:

$$B(\nu) = \frac{\sin \pi\nu/2\nu_{Nyq}}{\pi\nu/2\nu_{Nyq}}$$

- B(ν) is a broad function that reaches 0 at $2\nu_{\mathrm{Nyq}}$ and has the value of $2/\pi$ at ν_{Nyq}.

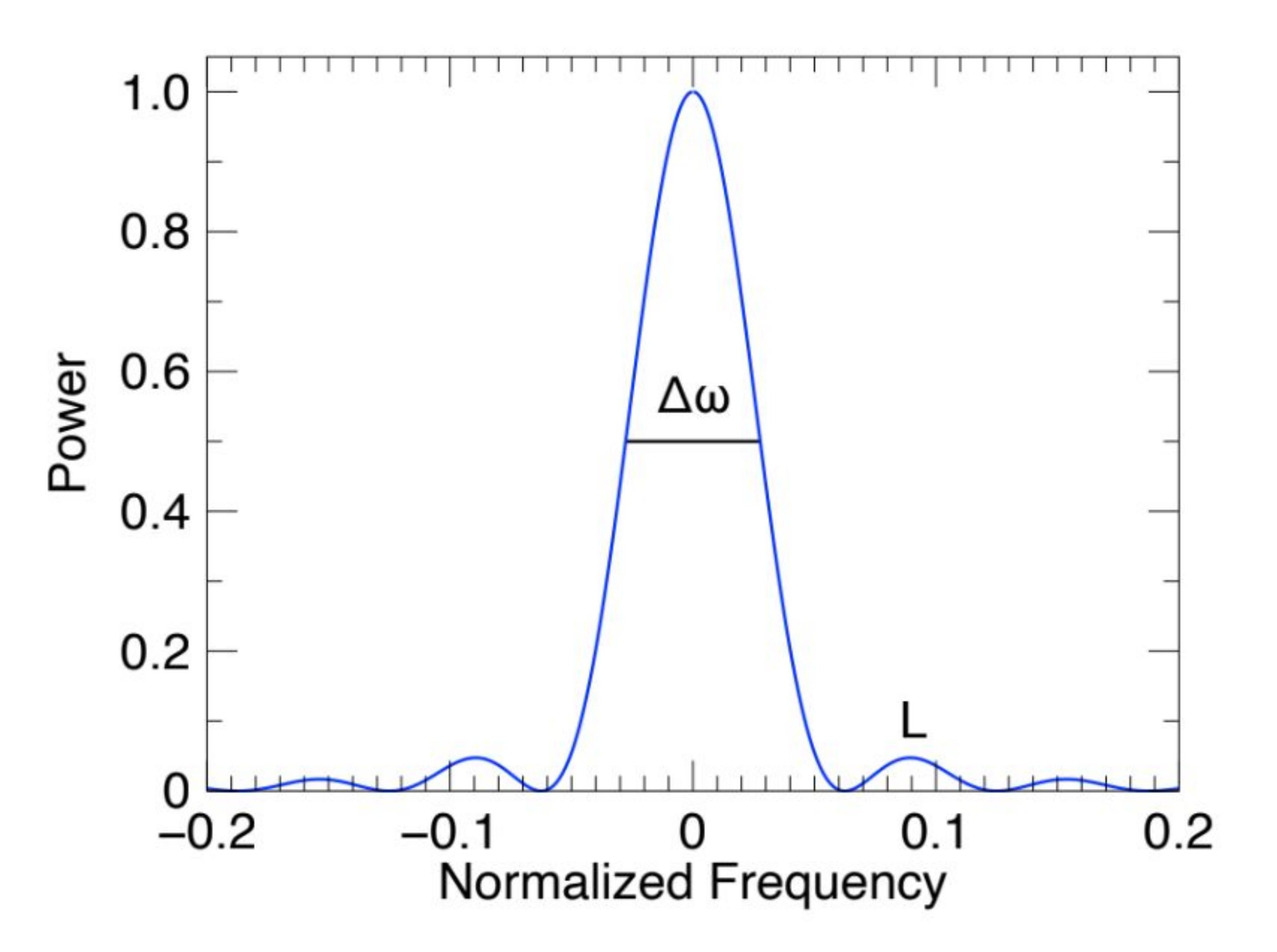
Window carpentry

- Having a longer observation reduces the width of the main window peak, but does not change the possible spillover effects.
- In some cases it can be advantageous to multiply the data by another window, not boxcar-shaped.
- This results in a loss of signal, as some data are multiplied by a factor less than unity, but there are advantages, depending on the chosen window.

- Many window functions with different characteristics have been designed and one can tailor them depending on what is needed.

Window carpentry

- The main features that identify a window in its PDS are: the width of the main peak Δω, the relative amplitude of the first side lobe L (expressed in decibels) and the slope of the decay of side lobes n:



Window carpentry

- The boxcar window is the one with the lowest $\Delta\omega$, but with alternative windows it is possible to obtain a significant reduction of the amplitude of the side lobes.

Window	$\Delta\omega$	L	n	Function
Boxcar	0.89	-13db	2	1
Hamming	1.36	-43db	2	$0.54 + 0.46 \cdot \cos(2\pi t)$
Hann	1.44	-32db	5	$0.5 \cdot [1 - cos(2\pi t)]$
Blackman	1.68	-58db	5	$0.42 + 0.5 \cdot \cos(2\pi t) + 0.08 \cdot \cos(4\pi t)$
Gaussian	1.55	-56db	2	$exp(-4.2 * x^2)$

Window carpentry

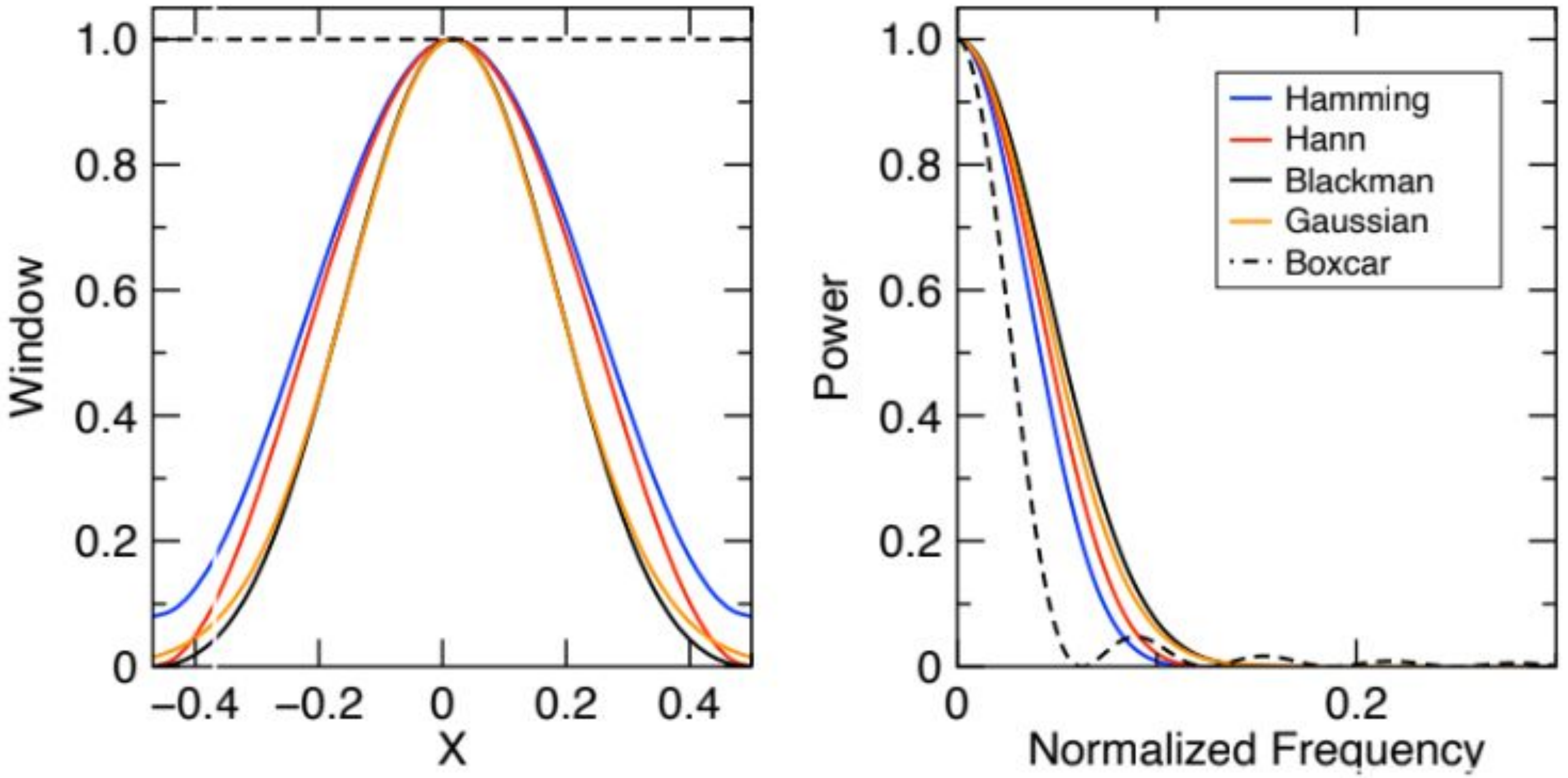


Fig. 8: Left panel: the shape of five windows: boxcar, Hamming, Hann, Blackman and Gaussian. Right panel: the corresponding PDS, corresponding to (half) the shape of the PDS of a sinusoidal signal. The sidelobes of all but the boxcar window are too low to be seen (see Tab. 2).

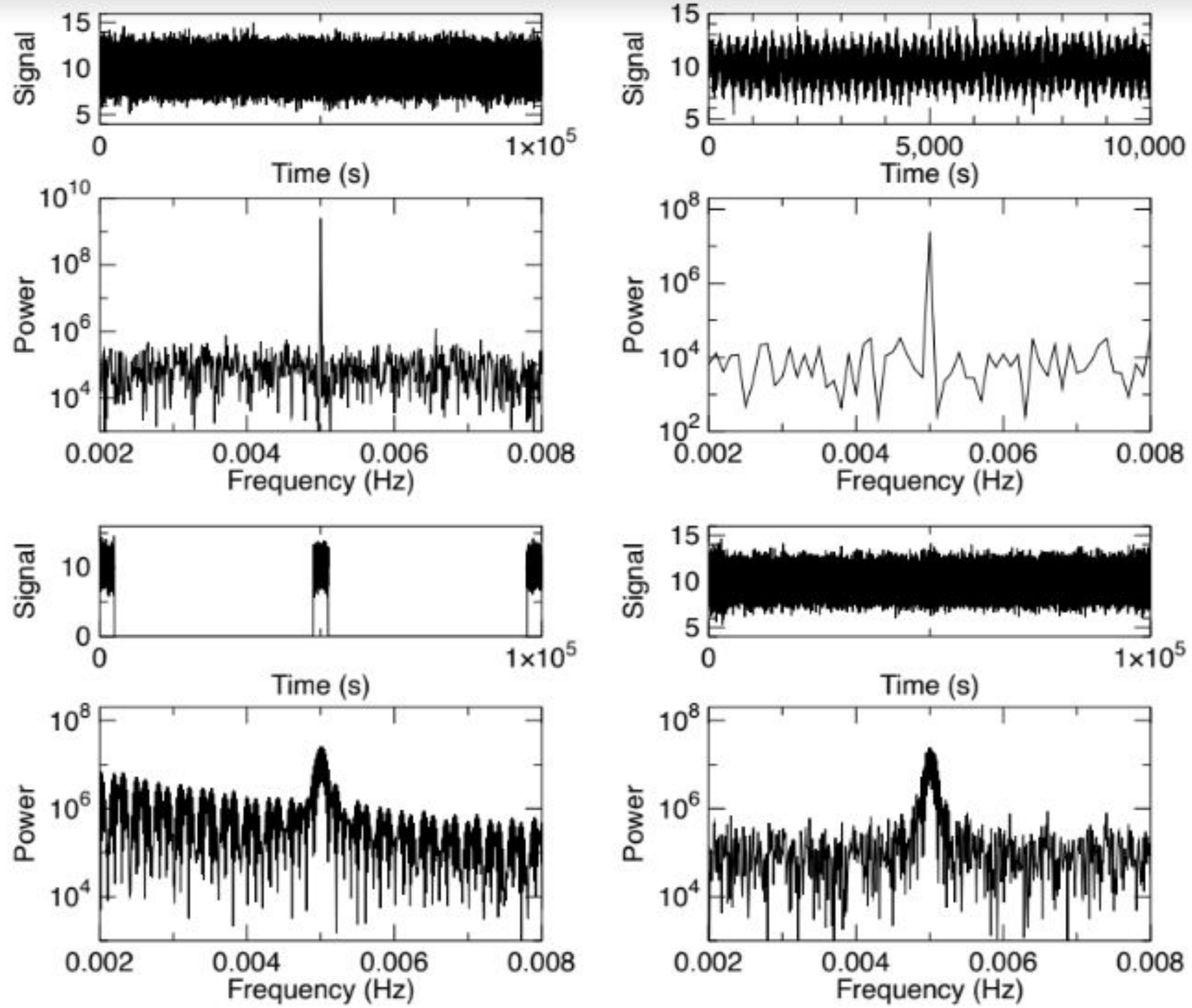


Fig. 9: Each of the four panels contains a signal (top) and its PDS (bottom). In all four, the signal consist of Gaussian noise plus a sinusoid at P=200 s, with 1-s binning. Top left: continuous exposure of 10^5 s. Top right: continuous exposure of 10^4 s. Bottom left: exposure of 10^4 s split into three intervals. The PDS is computed including the gaps as consisting of points at zero level. Bottom right: exposure of 10^4 s split into three intervals as in the previous case. The gaps are filled with Gaussian noise.

PSD normalization

- The FT is a linear transformation and the PSD is its squared modulus, the PSD then scales with the square of the intensity level of the signal.
- It is possible to normalize the PSD in different ways. One of the most common is the so-called *Leahy normalization*:

$$P_j^{Leahy} = \frac{2}{N_\gamma}|a_j|^2$$

- where N_γ is the total number of photons in the signal.
- If instead of counts we have fluxes, and noise is $N(0,\sigma^2)$, N_γ is substituted by $N_{data}\sigma^2$.

- This is also known as Classical or Schuster periodogram.

Leahy normalization

- This normalization leads to a known statistical distribution of signal power:
  - if the signal is dominated by fluctuations due to Poisson statistics and if N_γ is large, powers follow a χ^2 distribution with 2 degrees of freedom, $<P>=2$ and $Var(P)=4$.
- The reason is that the periodogram is the sum of the squares of the real and imaginary parts of the FT.
  - For a stochastic process, the latter are normally distributed, so the sum of their squares is distributed as a chi square with 2 degrees of freedom.
- For other noise distributions (Poisson, etc.), for large N, due to the central limit theorem the real and imaginary parts become still normal.

Bartlett's method and data gaps

- Periodograms are intrinsically very noisy.
- If the signal is divided into S segments and the resulting PSD are averaged and rebinned by a factor M, the powers will be distributed as a chi square with 2SM degrees of freedom scaled by 1/SM: therefore, the average power remains $\langle P \rangle = 2$, but the variance is now $Var(P) = 4/SM$.
  - The technique of dividing the time series into equal-duration intervals and averaging the corresponding PSD is called *Bartlett's method*.
- The reduction in time duration T increases the minimum frequency in the PSD $\nu_{min} = 1/T$.
- This method, in principle, also allows one to skip over (short) data gaps, which have dramatic effects on the PSD.

Auto and Cross-Correlation

- The Cross-correlation of two functions f(t) and g(t) is defined as:

$$C(\tau) = f \star g = \int_{-\infty}^{\infty} f^*(t) g(t+\tau) dt$$

- The result is a function of the "lag" τ introduced between the two functions, and is often used to estimate the similarity between two different time series, as a function of lag.
- If a common underlying process causes the time variation of intensity at two different electromagnetic bands with differential delays while propagating to the observer, then the cross correlation function of the two time series will exhibit a peak at the corresponding lag, namely the relative delay between the two bands.

Auto and Cross-Correlation

- The autocorrelation function is a special case where a function is correlated with itself, which would always show a peak at zero lag.
- Convolution is an operation akin to the Cross-correlation, but the function g(t) in the integrand is inverted to g(−t) before adding the shift.
-
- A few important properties of Cross-correlation include:

$$[f \star g](\tau) = [g^* \star f^*](-\tau)$$

$$[f \star g] \star [f \star g] = [f \star f] \star [g \star g]$$

$$g \star (f \otimes h) = [g \star f] \otimes h$$

$$F[f \star g] = F(f) \cdot F^*(g)$$

- where F represents the Fourier transform.

Cross-spectra, phase lag spectra, coherence

- Given two signals f(t) and g(t) and their respective FTs, F(ω) and G(ω), we define the cross spectrum as:

$$CS(\omega) = F(\omega) \cdot G^*(\omega)$$

- Analogous to the PSD and the autocorrelation, the cross spectrum between two signals is the Fourier transform of their cross-correlation (and vice-versa).

- In its essence, the cross spectrum of two signals at each frequency is a complex number whose argument represents the phase delay between the signals at that frequency.

Bispectrum

- The autocorrelation function may be thought of as the second order correlation:

$$c_2(\tau) = \langle f(t) f(t+\tau) \rangle$$

- where the angular brackets denote an ensemble average.

- Bispectrum is an extension of the above concept to triple correlations. The third order correlation function:

$$c_3(\tau_1, \tau_2) = \langle f(t) f(t+\tau_1) f(t+\tau_2) \rangle$$

- that allows one to define the Bispectrum:

$$B(\omega_1, \omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c_3(\tau_1, \tau_2) e^{-i(\omega_1 \tau_1 + \omega_2 \tau_2)} \, d\tau_1 \, d\tau_2$$

The Fourier Transform

- Useful notebooks:

1. FourierTransform
2. PowerSpectrum

Power Spectrum Statistics

- The first thing we need to know is the probability distribution of the noise power.

- We now assume that it is additive and independent of the frequency (i.e. noise is "white"):

$$P_j = P_{j,noise} + P_{j,signal}$$

- and a typical "null hypothesis" could be that the periodogram is consistent with pure noise.

- Let's remind us that If we have $x_k \equiv y_k + z_k$ and b_j and c_j are the FT of y_k and z_k, we have $a_j = b_j + c_j$. This does not hold for power spectra unless the signal are uncorrelated random noise:

$$|a_j|^2 = |b_j + c_j|^2 = |b_j|^2 + |c_j|^2 + \text{cross terms.}$$

Power Spectrum Statistics

- For a wide range of types of noise, $P_{j,noise}$ follows a χ^2 distribution with 2 degrees of freedom (but at the Nyquist frequency, where dof is 1).

- For other noise distributions (Poisson, etc.), for large N, due to the central limit theorem A_j and B_j become still normal.

- This suggests a simple consistency test: compute the standard deviation in each frequency bin and divide by the mean power. Results should be (if the hypotheses hold) close to 1.

- In practice one finds that noise powers are nearly always χ^2 distributed, not only for Poisson noise, but also for many other types of noise.

Power Spectrum Statistics

- With the Leahy normalization, the probability for $P_{j,noise}$ to exceed a given threshold is given by:

$$\mathrm{Prob}(P_{j,\mathrm{noise}} > P_{\mathrm{threshold}}) = Q(P_{\mathrm{threshold}}|2) \qquad (j = 1, N/2 - 1),$$

$$Q(\chi^2|\upsilon) \equiv \left[2^{\upsilon/2}\Gamma\left(\frac{\upsilon}{2}\right)\right]^{-1} \int_{\chi^2}^{\infty} t^{\frac{\upsilon}{2}-1} e^{-\frac{t}{2}}\, dt,$$

- where ν is the number of dof and Γ is the gamma function, the generalization of factorial to real and complex numbers.
  - Γ(n) = (n − 1)!
- For ν = 2, $Q(\chi^2 | 2) = \frac{1}{2} \int_{\chi^2}^{\infty} e^{-t/2} dt = e^{-\chi^2/2}$

Power Spectrum Statistics

- Therefore power spectra are unavoidably very noisy.
- Standard deviation of noise power is equal to the mean value ($\sigma_{P_j} = \langle P_j \rangle = 2$).

- More interesting, this cannot be improved increasing the number of data points (i.e. the length of the time series). This merely increases the number of powers.

- One can decrease the large variance rebinning the power spectrum and/or dividing the data in multiple segments of equal length. This of course degrade the frequency resolution.
- If the number of segments is large the power statistics tends to become Normal.

Power Spectrum Statistics

- Let's define confidence detection level as the power with only ε probability to be exceeded by noise.
- However, this holds for a single frequency. If your spectrum consists of N_{trial} (independent) frequencies, the confidence detection level decreases to take into account the multiple trials:

$$(1-\epsilon)^{N_{\rm trial}} \sim 1-\epsilon N_{\rm trial} \quad \text{for } \epsilon << 1$$

- In general if noise is not Poissonian or Gaussian we have correlation between the $P_{j,noise}$.
- Noise power spectrum will not be flat anymore.
- However, often noise powers still follow a χ^2 distribution with 2 dof, but with a different normalisation (in general depending on j).

The Likelihood for Periodograms

- The χ^2 distribution defines a sampling distribution or likelihood, i.e. the probability distribution of observing a given data set given some underlying (true, unknown) power spectrum.
- If we define a model power $S_j(\theta)$ at frequency ν_j, specified by a set of parameters θ, we can then compute the probability of having observed periodogram power P_j at that same frequency:

$$p(P_j|S_j(\theta)) = \frac{1}{S_j(\theta)} \exp\left(-P_j/S_j(\theta)\right) .$$

- The likelihood (also known as Whittle likelihood) for a periodogram over N/2 observed powers P_j is then defined as the product of individual probabilities for each frequency ν_j.

The Likelihood for Periodograms

- One generally (and equivalently), defines the logarithm of the likelihood as the sum of logarithm of all probabilities, such that:

$$\log(\mathscr{L}(\theta)) = \sum_{j=1}^{N/2} \log(p(P_j|S_j(\theta)))$$
$$= \sum_{j=1}^{N/2} \left(\log(S_j(\theta)) - \frac{P_j}{S_j(\theta)} \right) .$$

- A likelihood for averaged periodograms can be derived from the $\chi^2_{2LM}/2ML$ sampling distribution for periodograms averaged over L independent segments and M independent neighbouring frequencies by:

$$\mathscr{L}_{\mathrm{avg}}(\theta) = -2ML \sum_{j=1}^{N/2} \left\{ \frac{P_j}{S_j(\theta)} + \ln S_j(\theta) + \left(\frac{1}{ML} - 1 \right) \ln P_j + c(2ML) \right\} ,$$

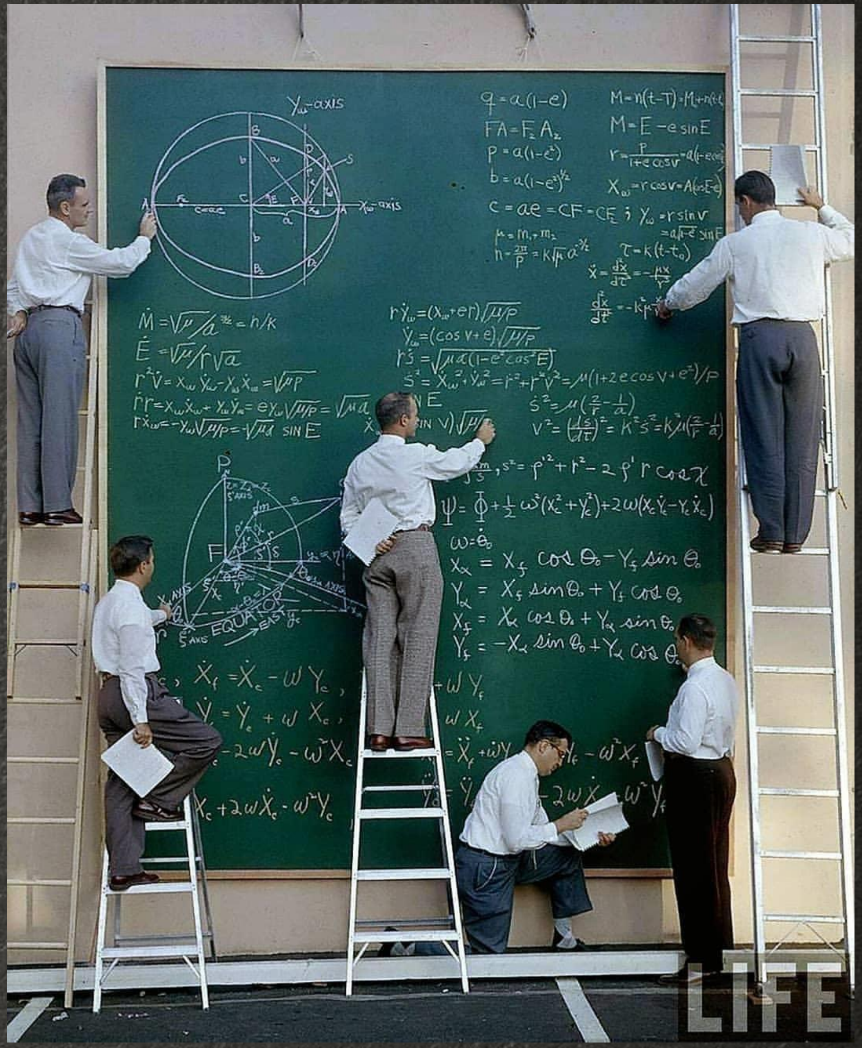
- where c(2ML) is a factor independent of P_j or S_j, and thus unimportant to the parameter estimation problem considered here.

THE “NOISE”!

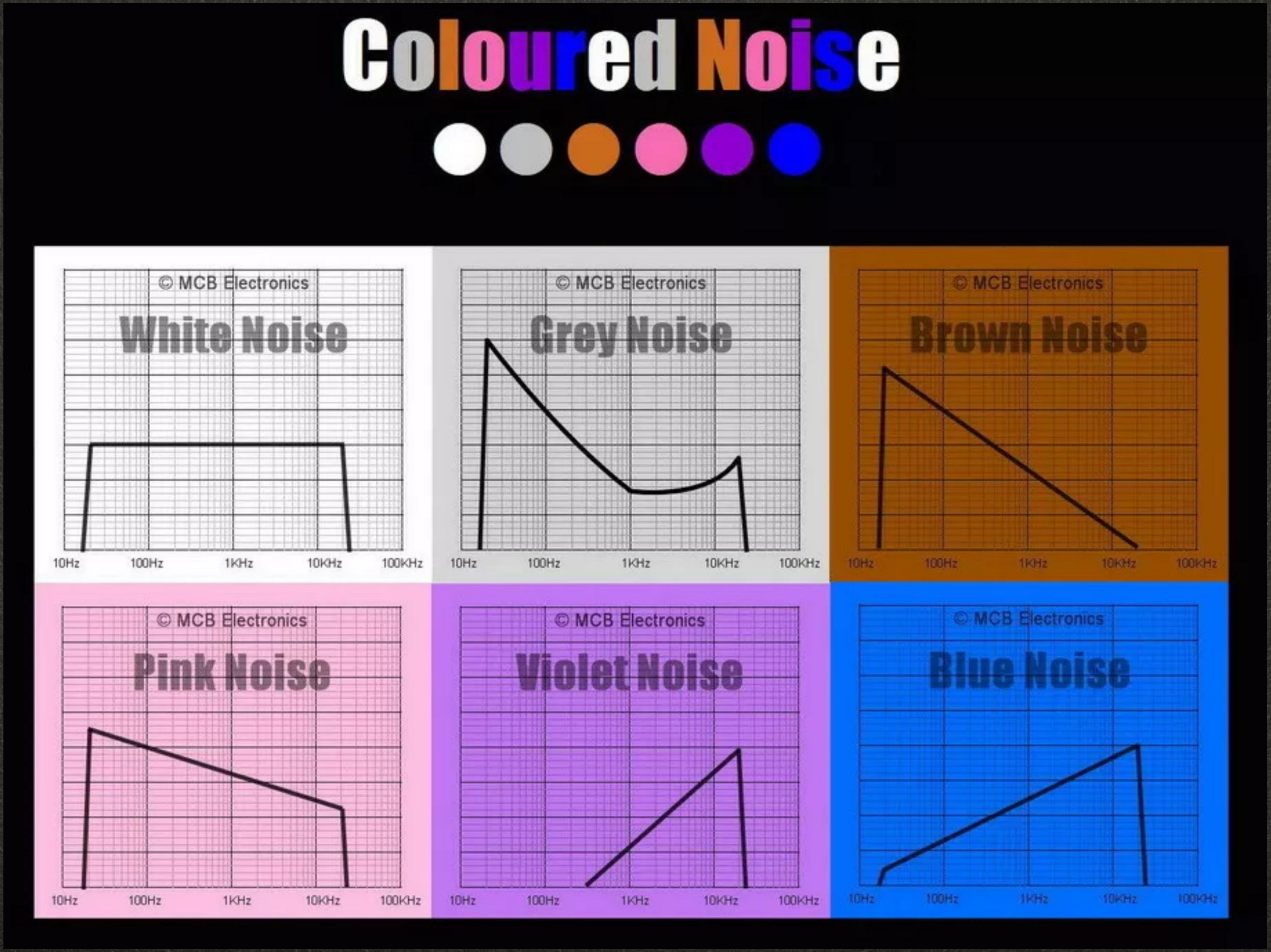
- Fourier analysis (and related techniques) has proven to be very effective in identifying periodic behaviours.

- In astrophysics however we often have to deal with phenomena too long for having a reliable coverage (decades or more) and with only approximately cyclical behaviours (quasi-periodicities).

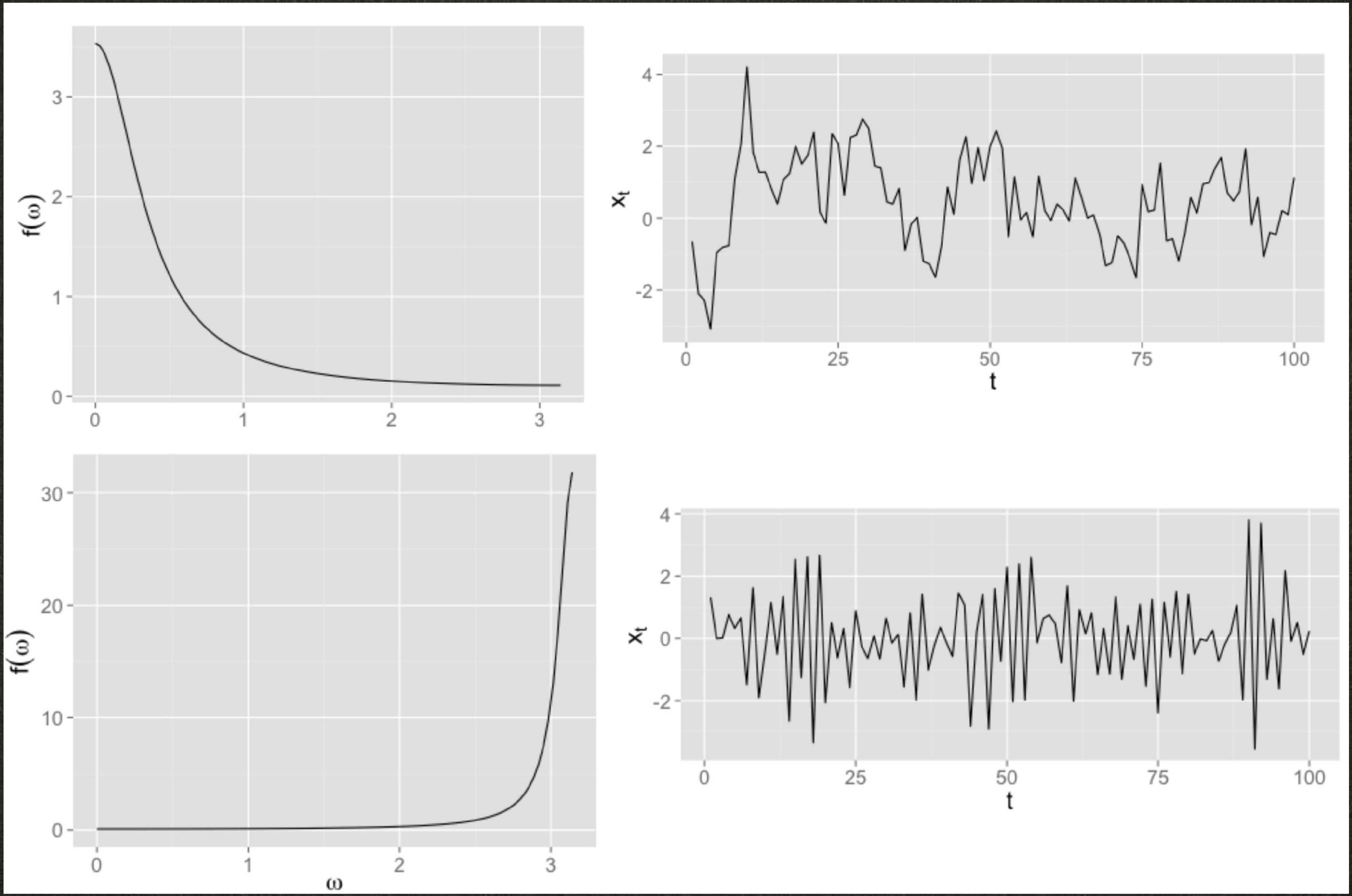
- More important, in order to compute the statistical significance of any possible periodicity, one needs to properly model the noise affecting a time series.



From a CD cover...



A couple of examples...



Summing up...

- Fourier analysis reveals nothing of the evolution in time, but rather reveals the variance of the signal at different frequencies.

- The classical periodogram is an estimator of the spectral density, i.e. the Fourier transform of the autocovariance function.

- Fourier analysis has restrictive assumptions: an infinitely long data of equally-spaced observations; homoscedastic Gaussian noise with purely periodic signal of sinusoidal shape.

- Formally periodogram statistics is simple to compute, but often is not applicable, and probabilities are difficult to infer.

In addition…

- The classical periodogram is not a good estimator, it is “inconsistent” because the number of parameters grows with the number of datapoints.

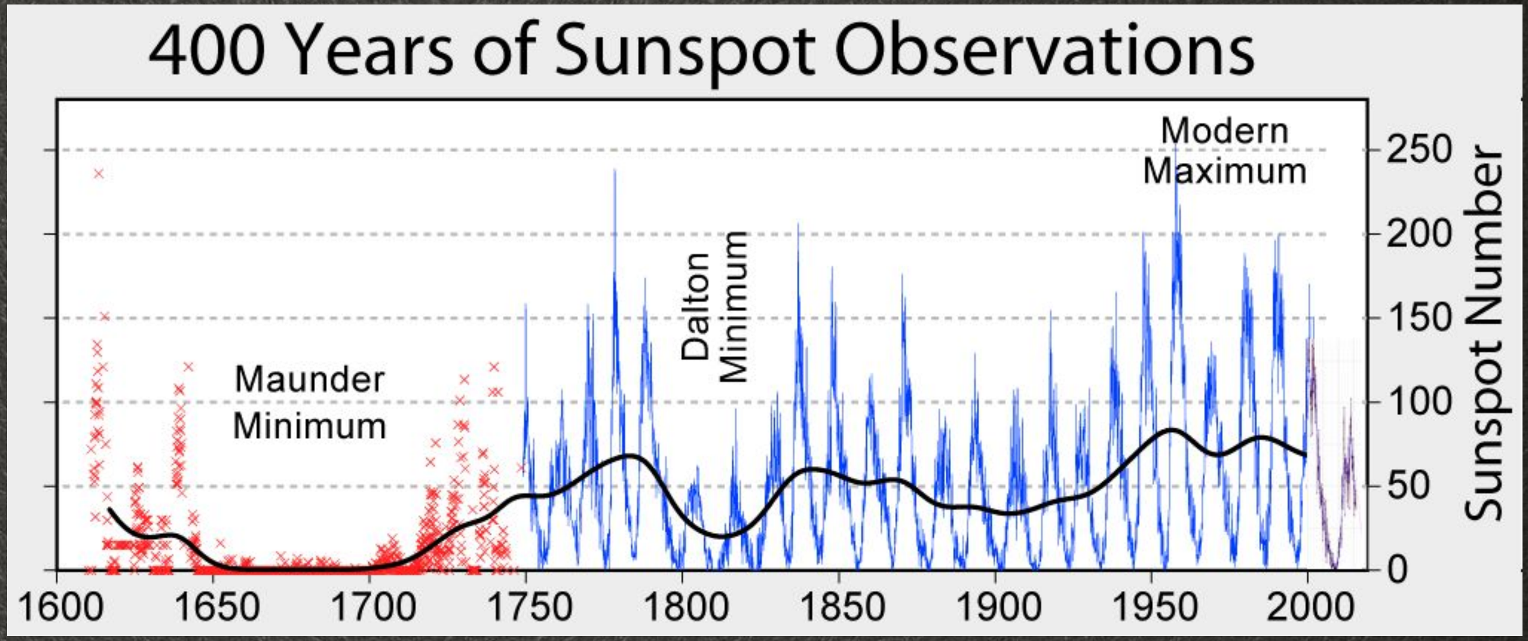
- The DFT and its probabilities depends on several strong assumptions that are rarely achieved in real astronomical data: evenly spaced data of infinite duration with a high sampling rate (Nyquist frequency), Gaussian noise, single frequency periodicity with sinusoidal shape and stationary behavior.

- Each of these constraints is violated in various astronomical problems. Data spacing may be affected by daily/monthly/orbital cycles. Periods may be comparable to the sampling time. Several periods may be present (e.g. helioseismology). Shape may be non-sinusoidal (e.g. elliptical orbits, eclipses, recurrent flares). Periods may not be constant (e.d. QPOs in accretion disks).

Science cases (simplified)

- Useful notebooks:

1. SunspotNumber
2. GX339QPO



REFERENCES AND DEEPENING

Basics of Fourier Analysis for High-Energy Astronomy

Tomaso M. Belloni * and Dipankar Bhattacharya

Abstract The analysis of time variability, whether fast variations on time scales well below the second or slow changes over years, is becoming more and more important in high-energy astronomy. Many sophisticated tools are available for data analysis and complex practical aspects are described in technical papers. Here, we present the basic concepts upon which all these techniques are based. It is intended as a condensed primer of Fourier analysis, dealing with fundamental aspects that can be examined in detailed elsewhere. It is not intended to be a presentation of detailed Fourier tools for data analysis, but the reader will find the theoretical basis to understand available analysis techniques.

FOURIER TECHNIQUES IN X-RAY TIMING

M. van der Klis,
EXOSAT Observatory,
Space Science Department of ESA,
ESTEC, Postbus 299,
2200 AG Noordwijk,
The Netherlands

ABSTRACT. Basic principles of Fourier techniques often used in X-ray time series analysis are reviewed. The relation between the discrete Fourier transform and the continuous Fourier transform is discussed to introduce the concepts of windowing and aliasing. The relation is derived between the power spectrum and the signal variance, including corrections for binning and dead time. The statistical properties of a noise power spectrum are discussed and related to the problems of detection (and setting upper limits) of broad and narrow features in the power spectrum. A "dependent trial" method is discussed to search power spectra consistently for many different types of signal simultaneously. Methods are compared to detect a sinusoidal signal, a case that is relevant in the context of X-ray pulsars.

Tomaso Belloni

Michiel Van der Klis

Dipankar Bhattacharya