Boring but important disclaimers:

- If you are not getting this from the GitHub repository or the associated Canvas page (e.g. CourseHero, Chegg etc.), you are probably getting the substandard version of these slides Don't pay money for those, because you can get the most updated version for free at

  `https://github.com/julianmak/academic-notes`

  The repository principally contains the compiled products rather than the source for size reasons.
- Associated Python code (as Jupyter notebooks mostly) will be held on the same repository. The source data however might be big, so I am going to be naughty and possibly just refer you to where you might get the data if that is the case (e.g. JRA-55 data). I know I should make properly reproducible binders etc., but I didn't...
- I do not claim the compiled products and/or code are completely mistake free (e.g. I know I don't write Pythonic code). Use the material however you like, but use it at your own risk.
- As said on the repository, I have tried to honestly use content that is self made, open source or explicitly open for fair use, and citations should be there. If however you are the copyright holder and you want the material taken down, please flag up the issue accordingly and I will happily try and swap out the relevant material.

OCES 3301 :

basic Data Analysis in ocean sciences

Session 8: fairly basic time-series analysis

Outline

(Just overview here; for actual content see Jupyter notebooks)

- Fourier analysis

  $\rightarrow$ background and theory

  $\rightarrow$ power spectrum

  $\rightarrow$ Fourier analysis as a filtering tool

Signal analysis

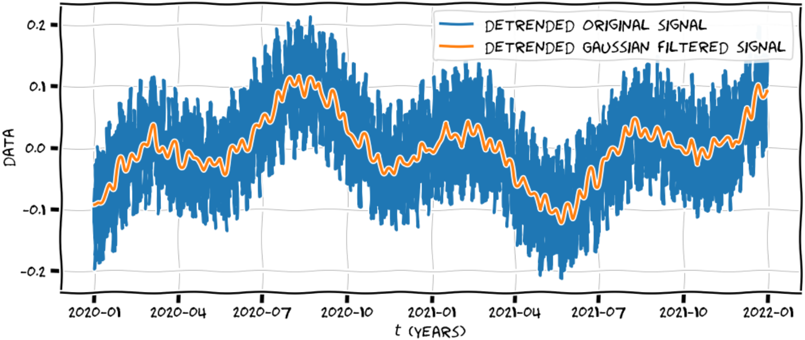


Figure: A detrended signal.

- low-passed signal still has some oscillations
- how to pick these out though?

  → can we say which period of oscillation is larger definitively?

Recall: waves

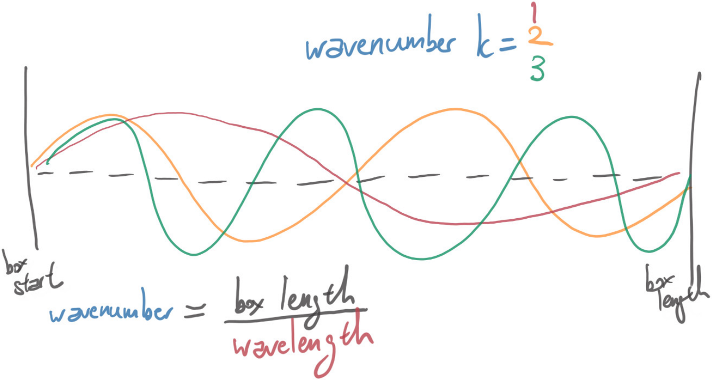


Figure: Schematic of wave features. Box length $L = 2\pi$ for simplicity.

- displacement η described by (could also be sine)

$$\eta \sim A\cos(2x), \qquad \eta \sim A\cos(1x), \qquad \eta \sim A\cos(3x)$$

Recall: waves

$$\gamma = \frac{v}{\lambda}, \qquad k = \frac{2\pi}{\lambda}$$

- γ the frequency (units: $\mathrm{s}^{-1} = \mathrm{Hz}$)

  $\rightarrow$ how quickly the wave oscillates
- $v = c_p$ the **phase** velocity

  $\rightarrow$ how fast the wave itself moves around
- λ the wavelength

  $\rightarrow$ how long the wave is
- k the wavenumber

  $\rightarrow$ intuitively how many waves can you fit in a box (so $k \sim \lambda^{-1}$)

  $\rightarrow$ does not necessarily have to be an integer

Fourier analysis

- claim (!!!): generically (!) a signal can **uniquely** be written as

$$
\begin{aligned}
f(t) &= a_0 + a_1\cos(t) + a_2\cos(2t) + \ldots + b_1\sin(t) + b_2\sin(2t) + \ldots \\
&= a_0 + \sum_{k=1}^{\infty} a_k\cos(kt) + \sum_{l=1}^{\infty} b_l\sin(lt), \qquad f \in L^2(P).
\end{aligned}
$$

- (k, l) are the wavenumbers

  $\rightarrow$ really angular frequencies here (since dealing with time)
- (a_k, b_l) are the amplitudes associated with the oscillations at (k, l)

magnitudes of (a_k, b_l) give you power of the oscillations at frequencies/wavenumbers at (k, l)

Fourier analysis: background

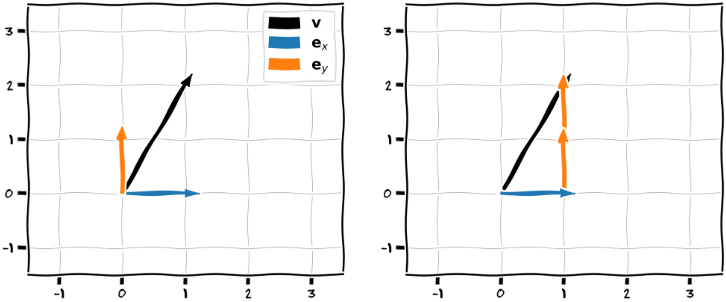


Figure: Demonstration of canonical basis in $\mathbb{R}^2$.

$$\mathbf{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 1\mathbf{e}_x + 2\mathbf{e}_y = 1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Fourier analysis: background

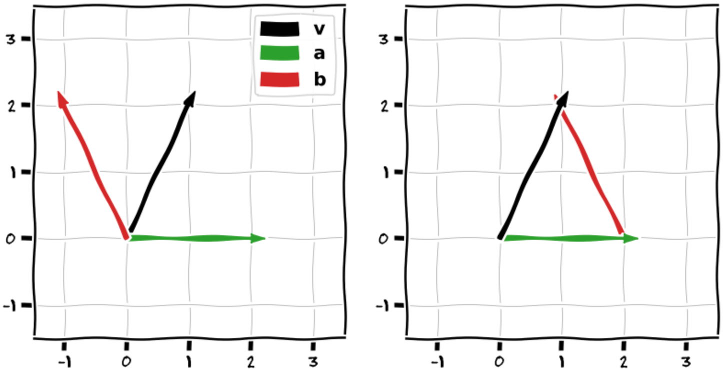


Figure: Demonstration of alternative basis in $\mathbb{R}^2$.

$$\mathbf{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 1 \begin{pmatrix} 2 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} -1 \\ 2 \end{pmatrix} = 1\mathbf{a} + 1\mathbf{b}$$

Fourier analysis: background

- cf. change of units (e.g. 1 inch $\approx$ 2.5 cm)
- cf. papillon vs. Schmetterling vs. فراشة
- cf. change of co-ordinates (e.g. PCA in 04)
- cf. uses different Lego combination to make the same thing
- cf. below?

- water, 35 L; carbon, 20 kg; ammonia, 4 L; lime, 1.5 kg; phosphorus, 800 g; salt, 250 g; saltpeter, 100 g; sulfur, 80 g; fluorine, 7.5 g; iron, 5 g; silicon, 3 g; trace amounts of 15 other elements

Figure: "Ed...ward..." (sorry not sorry)

Fourier analysis: background

A basis of (say) $\mathbb{R}^2$ is a set $\{\mathbf{v}_1, \mathbf{v}_2\}$, $v_i \in \mathbb{R}^2$ where

- the elements themselves are linear independent of each other, i.e.

$$\sum_i a_i \mathbf{v}_i = \mathbf{0} \qquad \Leftrightarrow \qquad a_i = 0 \quad \forall i$$

- the set spans $\mathbb{R}^2$, i.e.,

$$\forall \, \mathbf{u} \in \mathbb{R}^2, \; \exists \, \{b_i\} : \boldsymbol{u} = \sum_i b_i \boldsymbol{v}_i.$$

- note **uniqueness** follows from the two definitions
- no criteria about v_i being orthogonal to each other

  $\rightarrow$ elements of canonical basis $\{\mathbf{e}_i\}$ are mutually orthogonal and unit length (i.e. orthonormal)

Fourier analysis

Claim:

- $\{\cos(kt), \sin(lt)\}$ is a basis of the space of functions $L^2(P)$, where the space is equipped with the inner product

$$\langle f, g \rangle = \int_I f(t) g(t) \, \mathrm{d}t$$

- with respect to that inner product, the Fourier basis is an **orthogonal basis**, i.e.,

$$\int_I \sin(kt) \cos(lt) \, \mathrm{d}t = \int_I \sin(kt) \sin(lt) \, \mathrm{d}t = \int_I \cos(kt) \cos(lt) \, \mathrm{d}t = 0.$$

Fourier analysis

- a Fourier transform takes $f(t) \in L^2(P)$ into the co-ordinates (a_k, b_l)
  $\rightarrow$ e.g. by orthogonality,

$$\int_I f(t)\cos(t)\ \mathrm{d}t = \int_I \cos(t)\left[a_0 + \sum_{k=1}^{\infty} a_k \cos(kt) + \sum_{l=1}^{\infty} b_l \sin(lt)\right]\ \mathrm{d}t$$
$$= a_1 \int_I \cos^2(t)\ \mathrm{d}t,$$

  where we can in principle compute the integrals (numerically or otherwise)
  $\rightarrow f(t)$ is in time-domain or real space
  $\rightarrow (a_k, b_l)$ is in frequency-domain or spectral space

- `scipy.fft.fft/rfft` uses the **Fast Fourier Transform** to do it (Cooley–Tukey 1965 algorithm, although was actually known to Gauss 1876)

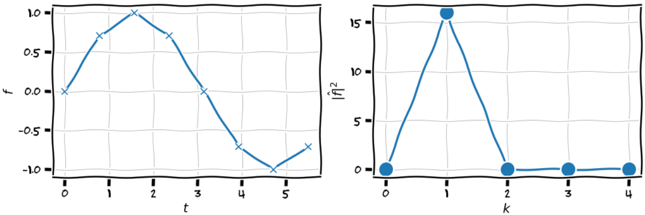
Fourier analysis

- a trivial example: taking $I = [0, 2\pi], f(t) = \sin(t)$
  $\rightarrow$ choice of I means wavenumbers $(k, l) \in \mathbb{Z}$
  $\rightarrow b_1 = 1$, everything else zero
- transform is (up to **normalisation**)

$$f(t) = (f(t_0), f(t_1), \ldots) \quad \leftrightarrow \quad \hat{f} \sim (a_0,\ a_1 + \mathrm{i}b_1,\ a_2 + \mathrm{i}b_2, \ldots),$$

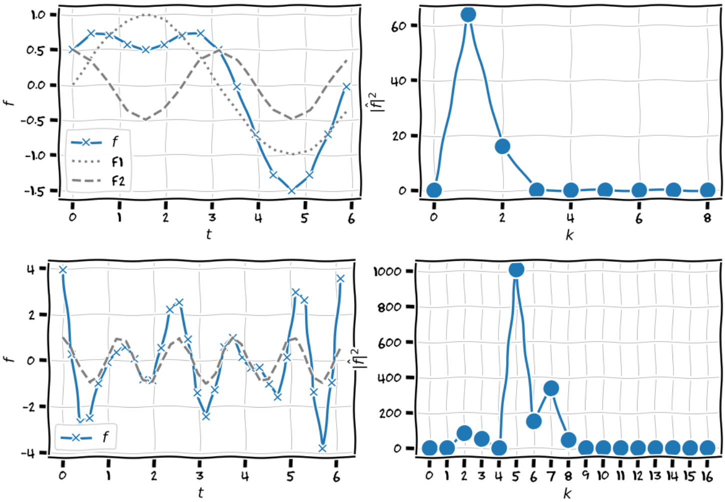
  where $\mathrm{i} = \sqrt{-1}$, and $(a_k, b_l) \in \mathbb{R}$
- power spectrum as a function of k is

$$|\hat{f}|^2 \sim (a_0^2,\ |a_1|^2 + |b_1|^2,\ |a_2|^2 + |b_2|^2, \ldots)$$



Fourier analysis

- still pure waves (sines and cosines) on $I = [0, 2\pi]$

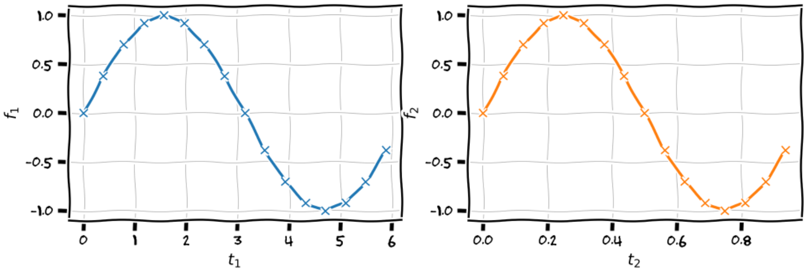


Fourier analysis: change of domain

- still pure waves (sines and cosines) but now on $I = [0, L]$
- take instead

$$\{\cos(kt), \sin(lt)\} \quad \leftrightarrow \quad \left\{\cos\left(\frac{2\pi k}{L}t\right), \sin\left(\frac{2\pi l}{L}t\right)\right\}$$

- below is when $L = 1$, showing $f_2(t) = \sin(2\pi t)$ (so $l = 1$)



Fourier analysis: change of domain

- still pure waves (sines and cosines) but now on $I = [0, L]$
- it's just a stretch/contraction, so just need to re-scale the definition of (k, l), as

$$k(\in \mathbb{Z}) \quad \rightarrow \quad \frac{2\pi k}{L}$$

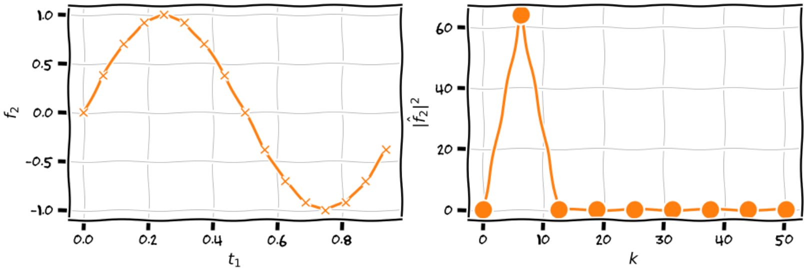


Figure: For $I = [0, 1]$. Compare with result on $I = [0, 2\pi]$ four slides ago.

Fourier analysis: change of units

Taking t to have units of seconds:

- k has units of radians per second (angular frequency here)
- $k = 2\pi\nu$, so frequency $\nu = k/(2\pi)$ and has units of per second, or **Hertz** (Hz)
- $\nu = 1/T$, so period $T = 2\pi/k$, here having units of seconds

- again, start with $k \in \mathbb{Z}$ and re-scale accordingly

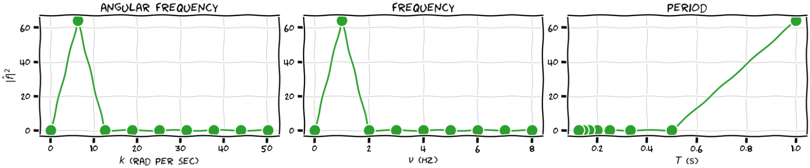
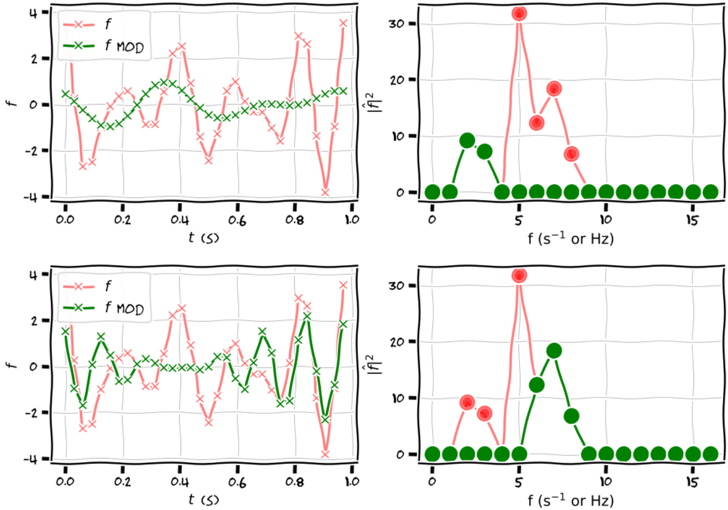


Figure: For $I = [0, 1]$, as previous slide, but in different units.

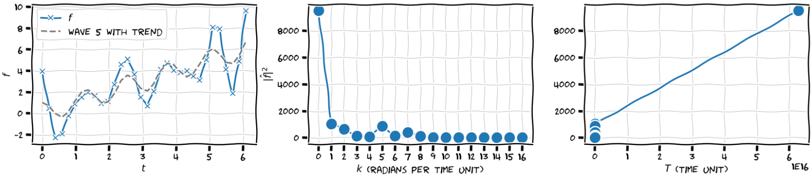
Fourier analysis: filtering

- modify the spectrum and do inverse Fourier transform



Things to be careful: signals with trends / non-zero

- a signal with some trend



$\rightarrow$ massive power in the $k = 0$ mode and in one with massive period

- non-zero mean so $a_0 \neq 0$, but this thing has no period

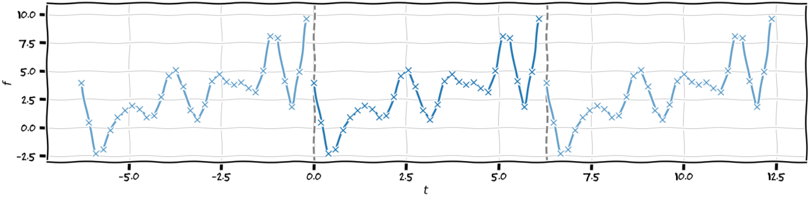
  $\rightarrow$ if you know there is a trend then detrend it

  $\rightarrow$ you could also get rid of the mean

  $\rightarrow$ or ignore / don't plot it, because the mean has no defined period of oscillation anyway

Things to be careful: non-periodic signal

- signal is not periodic (i.e. not strictly in $L^2(P)$)



- the FFT routines don't actually care, and will go ahead and return a spectrum

Things to be careful: non-periodic signal

- can force signal to be periodic through a window function

(cf. lec 07)

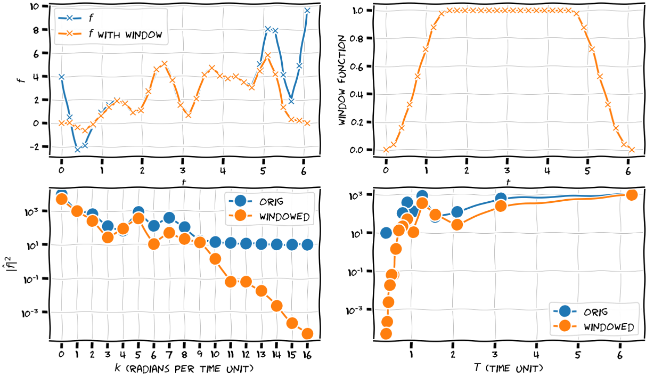


Figure: Example with Tukey window.

- different window functions looks different in real space, and have different responses in spectral space

Things to be careful: non-smooth signal

- Fourier transform actually still works

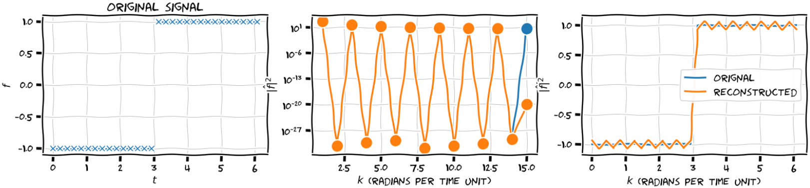


Figure: Classic example of square wave with Gibbs phenomenon.

- power is now everywhere and decays slowly with increasing wavenumber
- minor modification in spectrum leads to saw-tooth pattern

  $\rightarrow$ Gibbs phenomenon or Gibbs oscillations

Things to be careful: non-smooth signal

- signal with noise

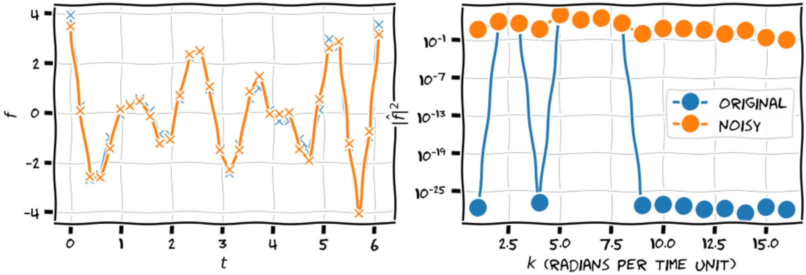


Figure: Power spectrum of signal, and signal with a bit of noise.

- power is now everywhere and decays slowly with increasing wavenumber

A real example: tides (see OCES 2003, lec 18)

Figure: High (or flood) and low (or ebb) tide at Tobermory, Isle of Mull, Scotland, using the pastel pink and red house as references. Modified images from `www.thechaoticscot.com` (left) and from myself (right).

A real example: tides (see OCES 2003, lec 18)

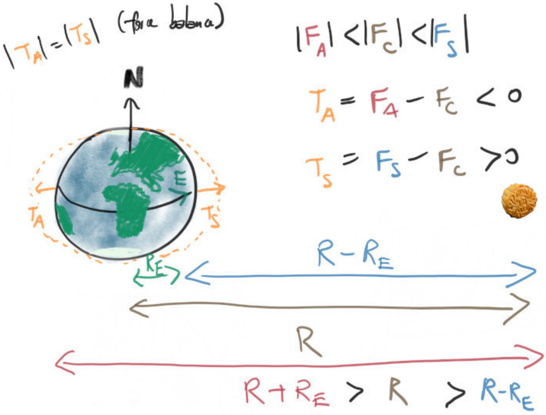


Figure: Schematic of tidal forcing by an astronomical body. Assume instantaneous response ("equilibrium theory"). No rotation is assumed here.

A real example: tides (see OCES 2003, lec 18)

symbol	period (in solar hrs)	rel. amp (to M_2)	name
M_2	12.42	1	principal lunar (semi-diurnal)
K_1	23.93	0.58	luni-solar (diurnal)
S_2	12.00	0.47	principal solar (semi-diurnal)
O_1	25.82	0.42	principal lunar (diurnal)
N_2	12.66	0.19	larger lunar elliptic (semi-diurnal)
$\vdots$	$\vdots$	$\vdots$	$\vdots$
Mf	327.85 ($\approx$ 14 days)	0.09	lunar fortnightly
Mm	661.30 ($\approx$ 28 days)	0.05	lunar monthly
SSa	4382.86	0.04	solar semi-annual

Table: Some sample tidal forcings sorted by relative amplitude to the M_2 tide (which is the largest forcing for Earth). Subset of Table 6.2 given in Wunsch (2015). The last few entries are weak and long term but they are there.

- M_2 and K_1 the dominant ones
  $\rightarrow$ usually do include these two in numerical models
- notice the periods are close to multiples of 12 hours

A real example: tides (see OCES 2003, lec 18)

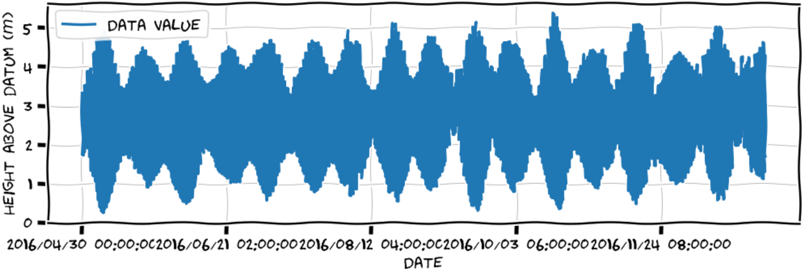


Figure: Data segment from BODC of sea level above datum at Tobermory.

- short-ish period, ignore trend with rising sea level
  $\rightarrow$ going to be ignoring $k = 0$ mode anyway

A real example: tides (see OCES 2003, lec 18)

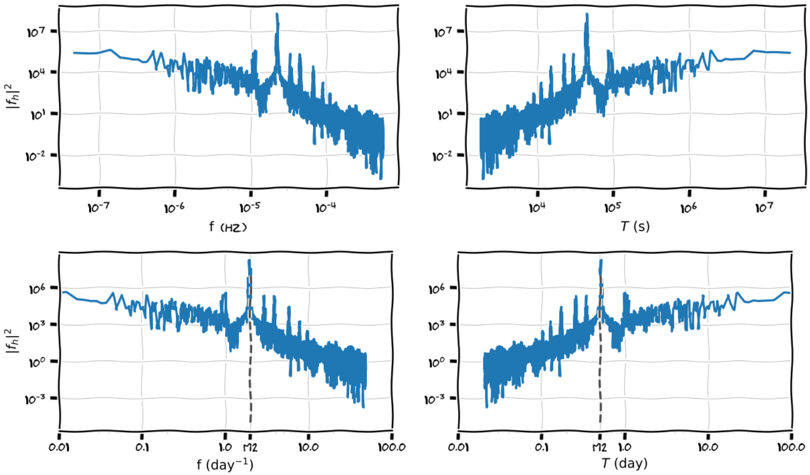


Figure: Power spectrum with respect to different quantities in different units, and denoting the M2 tide.

Jupyter notebook

go to 08 Jupyter notebook to get some code practise

- try something similar for the El-Niño 3.4 data

  $\rightarrow$ be careful of trends

  $\rightarrow$ be careful of units (time units is in **years**)

Note: none of the content I introduced in 'times series' are exclusive to 'time', and works just as well for 'space' too (see lec 09 and 10)