MPO624 Yueyang Lu

0. Why would a sensible person decompose or expand their data series into a sum of harmonic (sinusoidal) functions?

By doing that, we transfer the data into the frequency space, which can quantify the relative importance (power spectrum) of each constitutive sinusoidal wave in the total signal. We can extract the most significant signal from the data that is a collective result of many waves.

a.
On basin scale, Earth's atmosphere and ocean are ultimately driven by the (thermodynamic) solar heating and (dynamic) lunar forcing, both of which have a well-defined astronomical periodicity. The atmosphere and ocean will response to the imposed forces by generating variabilities like seasonal variability of surface temperature and tidal fluctuations. Fluid also have free oscillations on the rotational reference frame. 'Free' means there is no external forces to the object. This is a natural variability of fluid on the Earth.

b.
The existence of spectral peaks could improve the predictability because the peaks indicate dominate oscillations in the data.

1. What are the definitions and units of *wavelength*, *period*, *frequency*, *wavenumber*, *amplitude*, *phase*? What does *monochromatic* mean?

Wavelength $[m]$: the spatial period of a periodic wave, i.e., the distance between successive crests of a wave.
Period $[s]$: the temporal period of a periodic wave, i.e., the the time taken for one complete cycle of a wave.
Frequency $[s^{-1}]$: the number of full cycles per unit time.

Wavenumber $[m^{-1}]$: the number of full cycles per unit distance.
Amplitude: the measure of the change of a periodic variable over a single period. The unit is the same as the variable.

Phase $[rad]$: an angle representing the number of periods spanned by the argument of a periodic function. Suppose a function $F(t) = sin(2\pi t)$ with period $T = 1$. The phase is

$$\phi(t) = 2\pi \frac{t - t_0}{T} = 2\pi t \quad [rad].$$

Monochromatic describes a wave with single constant wavelength and frequency.

2. For a discrete data series V(t) = {Vi} with N values spaced dt apart, spanning the finite interval [0,T] with T = (N-1)dt, write down its Fourier decomposition:

a.

Assuming N is an odd number, the Fourier decomposition in sinusoidal terms is

$$\mathrm{V}_i = a_0 + \sum_{j=1} \left[a_j \cos(j\omega_0 \mathrm{t}_i) + b_j \sin(j\omega_0 \mathrm{t}_i) \right],$$

where $\omega_0 = 2\pi/T = 2\pi/((N-1)\mathrm{dt})$ and
$\mathrm{t}_i = i\,\mathrm{dt},\ i = 0, 1, .., N-1$, representing the N time steps.

The highest possible frequency (Nyquist frequency) is $f_{max} = 1/(2\,\mathrm{dt})$, determined by the spacing of data. This gives the maximum wave mode that can be represented on the discrete grid,

$$j_{max} = \frac{\omega_{max}}{\omega_0} = \frac{2\pi f_{max}}{\omega_0} = \frac{2\pi/(2\,\mathrm{dt})}{2\pi/T} = \frac{N-1}{2}.$$

The Fourier series is thus

$$\begin{aligned}\mathrm{V}_i &= a_0 + a_1\cos(1\omega_0 \mathrm{t}_i) + b_1\sin(1\omega_0 \mathrm{t}_i) + .. + a_{j_{max}}\cos(j_{max}\omega_0 \mathrm{t}_i) + b_{j_{max}}\sin(j_{max}\omega_0 \mathrm{t}_i) \\ &= a_0 + a_1\cos(1\frac{2\pi i}{N-1}) + b_1\sin(1\frac{2\pi i}{N-1}) + .. + a_{j_{max}}\cos(\pi i) + b_{j_{max}}\sin(\pi i)\,.\end{aligned}$$

The coefficient $b_0 = 0$ because the sine term for zero frequency is zero.

The phase at the Nyquist frequency is $j_{max}\omega_0 \mathrm{t}_i = \pi i,$ whose sine is always zero for $i = 0, .., N-1$. So it also has only one coefficient $a_{\frac{N-1}{2}}$.

The unknown coefficients are $a_0, a_1, .., a_{\frac{N-1}{2}}, b_1, \ldots b_{\frac{N-1}{2}-1}$, where the b_0 and $b_{\frac{N-1}{2}}$ terms are dropped out. Thus, the total number of Fourier terms is $1 + \dfrac{N-1}{2} + \dfrac{N-1}{2} - 1 = N - 1$.

Note that the number of known values of the series is also $N-1$ rather than N because $T = (N-1)\,\mathrm{dt}$ implicitly gives the periodicity $V_0 = V_{N-1}$.

b.

The discrete cosine transform is

$$\mathrm{V}_i = \sum_{j=0}^{(N-1)/2} A_j \cos(j\omega_0 \mathrm{t}_i + \varphi_j), \quad i = 0, .., N-1,$$

where

$$A_j = \sqrt{a_j^2 + b_j^2}, \quad \varphi_j = \tan^{-1}\frac{b_j}{a_j}, \quad j = 0, .., (N-1)/2.$$

Note that $A_0 = a_0$.

c.

The complex version of the decomposition is

$$V_k = \sum_{j=0}^{(N-1)/2} c_j e^{i \cdot j\omega_0 t_k}, \quad k = 0,..,N-1,$$ where i is an imaginary number and k is the index for time steps. Expanding the complex exponential function using Euler's formula, we obtain the relation

$$\begin{cases} a_j = \mathrm{Re}(c_j) \\ b_j = \mathrm{Re}(ic_j) = -\mathrm{Im}(c_j) \,. \end{cases}$$

d.

The power spectrum is $\mathrm{P}(\omega) = a^2(\omega) + b^2(\omega)$.

Using cosine only, we have $\mathrm{P}(\omega) = A^2(\omega)$.

For complex version, $\mathrm{P}(\omega) = c^2(\omega)$. c*c or |c|^2

3. Suppose you have 100 years of tropical rainfall data at a point, one value per day, from a model.

a.

The 40-50d frequency band is
$912.5\ \mathrm{cyc}/100\mathrm{y} - 730\ \mathrm{cyc}/100\mathrm{y}$.

The frequency of each Fourier component resolved in the data is from $1\ \mathrm{cyc}/100\mathrm{y}$ to $18250\ \mathrm{cyc}/100\mathrm{y}$, where the lowest and highest frequencies are determined by the maximum and minimum periods, $T_{max} = 100\mathrm{y}$ and $T_{min} = 2\mathrm{d}$, respectively.

So there are $912 - 730 + 1 = 183$ spectral power estimates falling in the 40-50 day band. The DOF in the band is 366.

b.

Suppose the segment has the length of x days, the width of the 40—50d band would be $x/40 - x/50 = 0.005x$ cyc/$x$days. The expected segment should give a 40—50d band width of 1 cyc/$x$days so that only one estimate falls in the band. So we have $x = 1/0.005 = 200\,\mathrm{d}$. In this case, the 40-50d band can be written as 5 cyc/200d $- 4$ cyc/200d.

There are $100 * 365/200 = 183$ segments from the data, i.e., 183 estimates falling in the 40—50d band. Same result with (a).

c.

From the table, we get the F-statistics $F_{.05}(366, \inf) = 1.12.$ That means the actual power needs to be 1.12 times larger than the AR1 power to reject the null hypothesis with 95% significance.

d.

In general, it would be easier to pass the significance test since the power in the band increases due to the effect of aliased noise.

Suppose the variance of daily rainfall is σ_d^2. The extra 'aliased noise' variance, $4\sigma_d^2$, is uniformly distributed over the spectrum whose frequency is from 1 cyc/100y to 18250 cyc/100y.

there is no guarantee it is uniform!!

So the total 'aliased' variance added to the 40-50d band is

$$\frac{4\sigma_d^2}{18250} \times 183 = 0.04\sigma_d^2.$$