1. **Why** would a sensible person decompose or expand their data series into a sum of harmonic (sinusoidal) functions? Hint: This is to make you read my "crashcourse" writeup.

Well I think that a person would expand their data into a sum of harmonic functions is to help discover trends, oscillation patterns, diurnal, yearly (etc.) cycles. By decomposing the data it attempts to identify patterns or cycles in a time series data set which has already been normalized. https://www.investopedia.com/terms/f/fourieranalysis.asp

- a. List some reasons why we might we expect or hope to find such signals in the atmosphere and ocean. Be sure to use both the words forced and free in your answer, and explain why harmonic forcing is expected.

We would expect to find these signals because a lot of earth processes are based on seasonality, diurnal cycles, yearly cycles. – we do live on a rotating planet, and I think that it is not so surprising that we do see such frequent harmonic forcing in the Earth Sciences. We would hope to learn about frequency and timing of these cycles to be able to study changes in our earth system (i.e skipping the next glacial period).

- b. Explain the link between finding a statistically significant periodicity (a spectral peak) and predictability.

A spectral beak might tell us about the processes working together to generate the things we observe. The predictability depends on the periodicity of those peaks. (i.e the stock market example shows that if there is no strong periodicity to those peaks there can be no predictability, hence, why it is very difficult to predict stock market trends. On the contrary, it is very easy to predict diurnal temperature cycles at a specific latitude because there are real spectral peaks with periodicity. (I think)

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

Wavelength : It is the distance between consecutive corresponding points of the same phase on the wave. Units of length such as nm or μm.

Period : is the measure of time it takes for one wave cycle to complete units of time (s)

Frequency : Is the inverse of period with units of 1/t. It is how many cycles occur in a second (hz)

Wavenumber : is the spatial frequency of the wave measured in cycles per unit distance or radians

Amplitude : is a measure of its change in a single period. Units of oscillating variable

Phase : is the fraction of a complete cycle corresponding to an offset in the displacement from a specified reference point at time t = 0

Monochromatic : refers to a wave with a single frequency (i.e. Sine wave)

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

   a. Write it as a sum of cosine terms (with coefficients a0, a1, a2, a3...) and sine terms (with coefficients b1, b2, b3...), for the set of frequencies $\omega = \{0,1,2,3...\}$ times the lowest possible frequency or **fundamental harmonic** $\omega_0 = (2\pi/T)$. Follow equations 8.22-8.23 in the [Martinson book, p266](). The lowest or fundamental frequency ω_0 is also called the **bandwidth**, since it discretizes the frequency domain into finite bins or bands (a good use for a bar graph). After the ... where you stop enumerating the many frequencies after 0,1,2,..., make your expression end with explicit terms at the **highest** possible frequency, called the **Nyquist** frequency. Why is there no b coefficient for zero frequency (b0=0)? Why does the Nyquist frequency also have only one coefficient? With these two special zeros, show that N is how many Fourier coefficients it takes to express this series of N values

   see attached

   b. Write it again as a **discrete cosine transform**, using cosines only, but with phase offsets in each frequency (that is, with *amplitude A and phase φ* rather than *sine and cosine* components in each frequency). How are the A and φ coefficients related to a and b coefficients from part a? See Eq. (8.11) on p259 of the Martinson chapter linked above.

   See attached

   c. Write the complex version of the decomposition: $V = \sum_j c_j \exp(i\omega_0 t\cdot j)$. Show how a_j and b_j relate to the real and imaginary parts of complex c_j. This form is in [Wikipedia]().

   See attached

   d. Write the power spectrum $P(\omega)$ in terms of a and b coefficients; and also in terms of A and φ; and also in terms of the complex c.

   see attached

3. Suppose you have 100 years of tropical rainfall data at a point, one value per day, from a model. You want to ask if the model has the Madden-Julian oscillation, which Madden and Julian (1971) detected in nature: a statistically

$$F(t) = A_0 + \sum_{n=1}^{N} A_n \sin(w_n t + \phi_n)$$

$$A_n \sin(w_n t + \phi_n) = a_n \cos(w_n t) + b_n \sin(w_n t)$$

$W_0 = \frac{2\pi}{T}$ $T = N-1$ $N = [0, T]$ $w = \{1, 2, 3, 4 \dots\}$

$$\left.\begin{array}{l} a_0 \cos 0 \, w_t \cdot w = a_0 \cancel{\cos 0}^{1} \\ b_0 \sin 0 \, w_t \cdot w = b_0 \cancel{\sin 0}^{0} \end{array}\right\} a_0$$

$a_1 \cos(1 \cdot \frac{2\pi}{T} \cdot w_1) = a_1 \cos(\frac{2\pi}{T})$ $w = 1$

$b_1 \sin(1 \cdot \frac{2\pi}{T} \cdot w_1) = b_1 \sin(\frac{2\pi}{T})$ $W_0 = \frac{2\pi}{T}$

$a_2 \cos(2 \cdot \frac{2\pi}{T} \cdot 2) = a_2 \cos \frac{8\pi}{T}$ $W = 2$

$b_2 \sin(2 \cdot 2\pi/T \cdot 2) = b_2 \sin 8\pi/T$ $W_0 = \frac{2\pi}{T}$

$a_3 \cos(3 \cdot \frac{2\pi}{T} \cdot 3) = a_3 \cos \frac{18\pi}{T}$ $W = 3$

$b_3 \sin(3 \cdot \frac{2\pi}{T} \cdot 3)$ $b_3 \sin 18\pi/T$ $W_0 = \frac{2\pi}{T}$

This would only gives cos values because $\sin(2\pi)$ variations $= 0$ and $\cos(2\pi) = 1$

If the nyquist frequency is the highest frequency which you can represent sampling rate. divide sampling rate by 2 if highest frequency is 2π would Nyquist be π?

$\frac{1}{2}(2\pi)$??

$A\cos(\omega t + \varphi) = A\cos(\varphi)\cos(\omega t) - A\sin(\varphi)\sin(\omega t)$

φ = fixed angle $\Rightarrow A\cos(\omega t + \varphi) = a\cos(\omega t) - b\sin(\omega t)$

$a = A\cos(\varphi)$ and $b = A\sin(\varphi)$

$\varphi = \tan^{-1}\frac{b}{a}$ $\quad A = (a^2 + b^2)^{1/2}$

$\frac{\sin(\varphi)}{\cos(\varphi)} = \tan\varphi = \frac{b}{a}$

$\varphi = \tan^{-1}\frac{b}{a}$ 4 $\quad a^2 + b^2 = A^2[\cos^2(\varphi) + \sin^2(\varphi)]$

$\sin^2 x + \cos^2 x = 1 \Rightarrow A = (a^2 + b^2)^{1/2}$

$A_0\cos((0\omega_t \cdot \omega_t) + \varphi) + A_1\cos(1\omega_t \cdot \omega + \varphi) + A_2\cos 2(2\omega_t\,\omega + \varphi)$

$A_0\cdot \overset{1}{\cancel{\cos}}(0) + \cos(\varphi) + A_1\cos(1\cdot\frac{2\pi}{T}\cdot 1 + \varphi) + A_2\cos(2\omega_t\cdot 2 + \varphi)$

$\Rightarrow A\cos(\varphi) + A_1\cos(\varphi)\cos(\frac{2\pi}{T}) + A_1\sin(\varphi)\sin(\frac{2\pi}{T})$ ...

$A_2\cos(\varphi)\cos(2\cdot\frac{2\pi}{T}\cdot 2) - A_2\sin(\varphi)\sin(2\cdot\frac{2\pi}{T}\cdot 2)$

$$\cos(wt) = \frac{1}{2} e^{iwt} + e^{-iwt}$$
$$\sin(wt) = \frac{1}{2} e^{iwt} + e^{-iwt}$$
EULERS $e^{i\theta} = \cos\theta + i\sin\theta$

Complex $F(t) = \sum_{n=-N}^{+N} C_n e^{i w_0 t \cdot n}$ where $w_0 = \frac{2\pi}{T}$

where $C_n = \frac{b_n - i a_n}{2}$ $\quad -C_n = \frac{b_n + i a_n}{2}$

$C_0 = \frac{a_0}{2}$

wikipedia breakdown $x(K) = \sum_{n=0}^{N-1} x(n) \cos\left(\frac{2\pi K n}{N}\right)$

$$- i \sum_{n=0}^{N-1} x(n) \sin\left(\frac{2\pi K n}{N}\right)$$

when $x=0$ $\quad \sum_{n=0}^{N-1} x(n) \cos\left(\frac{2\pi 0 n}{N}\right) + i \sum_{n=0}^{N-1} x(n) \sin \frac{2\pi (0 n)}{N}$

$$= \sum_{n=0}^{N-1} x(n) + i \sum_{n=0}^{N-1} 0$$

~~Power spectrum = Real$(x(K))^2$ + Im$(x(K))^2$~~

okay so $C_0 = \frac{a_0}{2}$

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \left(\frac{1}{2} e^{i\theta} + \frac{1}{2} e^{-i\theta} \right) \right) + b_n \left(-\frac{1}{2} i e^{i\theta} + \frac{1}{2} i e^{-i\theta} \right))$$

$\Theta_1 = w_0 \cdot w^{1} \cdot t$

$$\frac{a_0}{2} + a_1 \left(\frac{1}{2} e^{i\theta_1} + \frac{1}{2} e^{-i\theta_1} \right) + b_1 \left(-\frac{1}{2} i e^{i\theta_1} + \frac{1}{2} i e^{-i\theta_1} \right)$$

$$+ a_2 \left(\frac{1}{2} e^{i\theta_2} + \frac{1}{2} e^{-i\theta_2} \right) + b_1 \left(-\frac{1}{2} i e^{i\theta_2} + \frac{1}{2} i e^{-i\theta_2} \right)$$

...

$\theta_2 = w_0 \cdot w_2 \cdot t$

$$P = \lim_{T \to \infty} \frac{1}{T} \int_0^T |x(t)|^2 dt$$

Fourier transform $\hat{x}(w) = \frac{1}{\sqrt{T}} \int_0^T x(t) e^{-iwt} dt$

do I need to use parsevals theorem?

$$A(x) = \sum_{n=\infty}^{\infty} a_n e^{inx}$$

$$B(x) = \sum_{n=\infty}^{\infty} b_n e^{inx}$$

$$S_{xx}(w) = \lim_{T \to \infty} E\left[|\hat{x}(w)|^2 \right]$$

where E is the $E\left[\frac{1}{T} \int_0^T x(t) e^{iwt} dt \int_0^T x(t') e^{-iwt} dt' \right] = \phi$

$$= \frac{1}{T} \int_0^T \int_0^T E\left[x(t)\, x(t') e^{(iwt - t')} \right] dt\, dt'$$

signifficant (above an AR1 red noise null hypothesis) spectral peak, *in the frequency band with 40-50 day periods*.

a. One simplest approach is to call numpy.fft() or scipy.periodogram() on the whole 100-year series. For this series with T = 100y and dt = 1d, **how many spectral power estimates fall in the 40-50 day band?** If you smooth the spectrum by averaging all these frequencies in the 40-50 day band together, your estimate for the power in this band with have twice that many degrees of freedom (DOFs), and won't have to be so extremely tall to pass the F-test. Remember, for just 2 DOFs, the F-test requires a factor of 10 (!) in variance for a peak or trough to be 95% significant.

see attached

b. Another approach is to chop the series into segments, take the power spectrum of each segment, and then average those. **How short a segment will have exactly 1 power estimate in the 40-50 day frequency band?** How many such segments can you make from 100 years of data? Your power estimate in this spectral band will have twice that many degrees of freedom (DOFs). **How does that compare to your answer from a.?** Hopefully the same?

See attached

c. Now you want to test if the power in the 40-50 day band is "significanttly" higher than red noise (an AR1 autogressive "red noise", your **null hypothesis**, whose rejection would support the hypothesis that the MJO is present). The test for comparing two variances is called the **F test**. Look it up and study its use (ens of Crashcourse has a link). **For the DOFs from a. and b., how large a ratio of your actual power to the AR1 "null hypothesis" power would have p < 0.05 (95% significance level)?**

Using the F test table it shows for 25% enhanced variance we need a ratio of 1.2539

d. Suppose your data were not actually daily rainfall, but hourly rainfall amounts, subsampled once per day. If the variance of hourly rainfall is 5x bigger than the variance of daily rainfall, the extra 400% of variance will be misinterpreted as low frequencies, or **aliased** into the range of your spectrum (between the bandwidth and the Nyqyuist frequency), as my crashcourse notes explain. If you assume that this extra 400% of "aliased sampling noise" variance is distributed over your spectrum uniformly over the frequency bins, what is its effect on the significance problem from c.? Discuss in sensible vocabulary terms; calculate something if you can.

If we have an increase in 400% variance we will not be able to detect the 25% enhanced variance. That increases the noise to signal ratio 15 fold. The aliased signal variance >>> signal enhancement

3a.

40 day frequency occurs ~9x a year

$\frac{40}{365} = 9.125 \times 100$ years $= 912.5$

50 day frequency occurs ~7.3 x a year

$\frac{50}{365} = 7.3 \times 100$ years $= 730$

$912.5 - 730 = 182.5$ DOf $= 182.5 \times 2 = 365$ 1/day!

b. lowest frequency $= \frac{1}{40} - \frac{1}{50} = .025 - .02 = .005$

.005 hz to get .005 hz we find 200 days is th correct cycle

$\frac{5}{200} = \frac{1}{40}$ and $\frac{4}{200} = \frac{1}{50}$ gives us .005 ✓

a 200 day cycle occurs 183.5 x in 100 years

$\frac{36500}{200} = 183.5$ same answer as a ✓