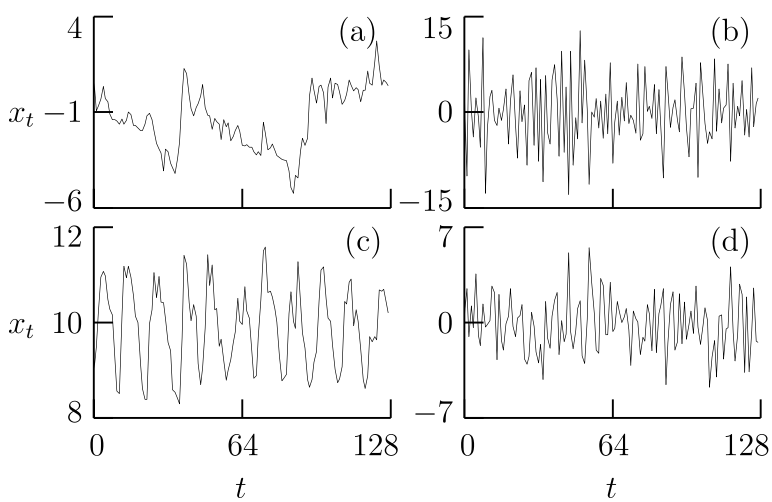
Introduction to Spectral Analysis

Don Percival, Applied Physics Lab, University of Washington

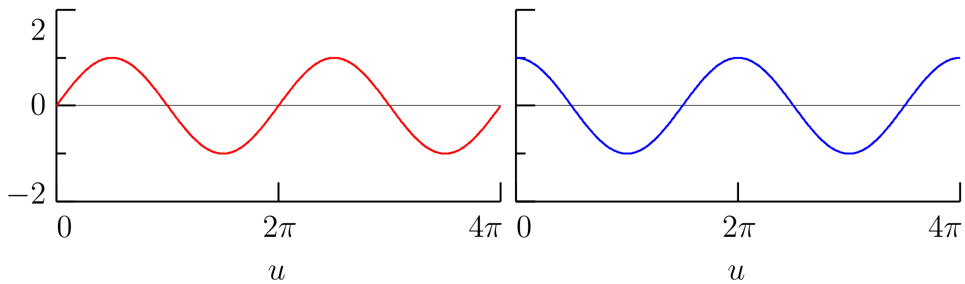
- Q: what is spectral analysis?
- one of the most widely used methods for data analysis in geophysics, oceanography, atmospheric science, astronomy, engineering (all types), ...
- method is used with time series
- let x_t denote value of time series at time t; for example, $x_{10} = 43^\circ$ = temperature at GHS at 8AM on 10th day of 2003
- four examples of time series $x_1, x_2, \ldots, x_{127}, x_{128}$



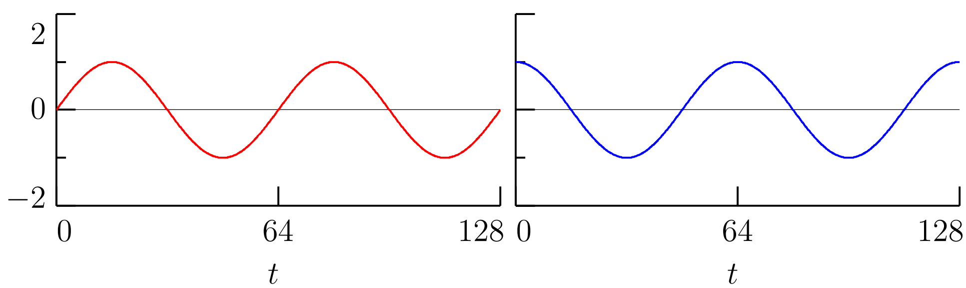
- Q: how would you describe these 4 series?
- spectral analysis describes x_t's by comparing them to sines and cosines

Sines and Cosines: I

- Q: what do sines and cosines have to do with time series?
- plots of sin(u) and cos(u) versus u as u goes from 0 to 4π



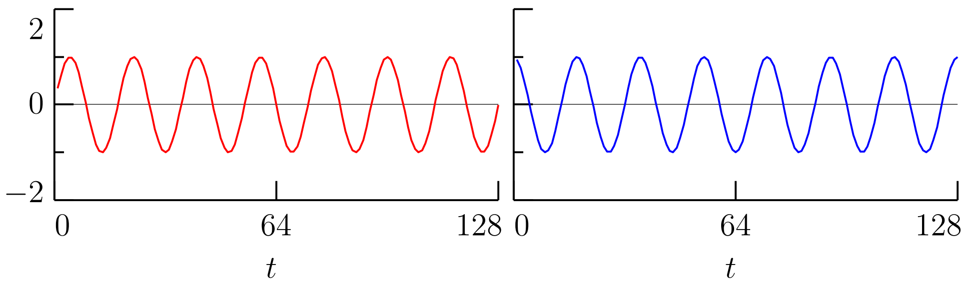
- let $u = 2\pi \frac{2}{128} t$ for $t = 1, 2, \ldots, 128$
- plots of sin$(2\pi \frac{2}{128} t)$ and cos$(2\pi \frac{2}{128} t)$ versus t



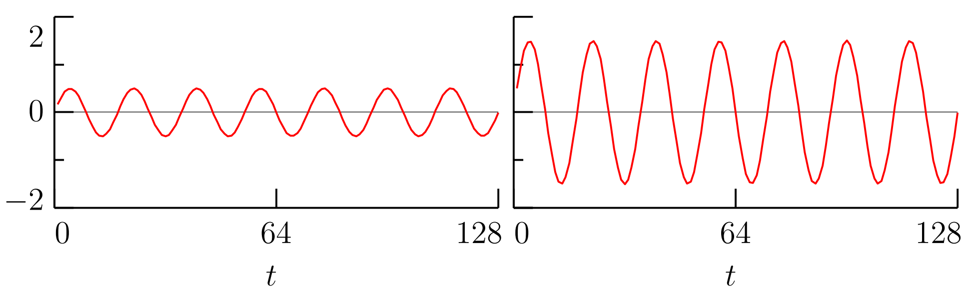
- can regard above as artificial time series: $s_t = \sin(2\pi \frac{2}{128} t)$ etc
- can interpret $\frac{2}{128}$ as '2 cycles over time span of 128'

Sines and Cosines: II

- now let $u = 2\pi\frac{7}{128}t$ for $t = 1, 2, \ldots, 128$
- plots of $\sin(2\pi\frac{7}{128}t)$ and $\cos(2\pi\frac{7}{128}t)$ versus t

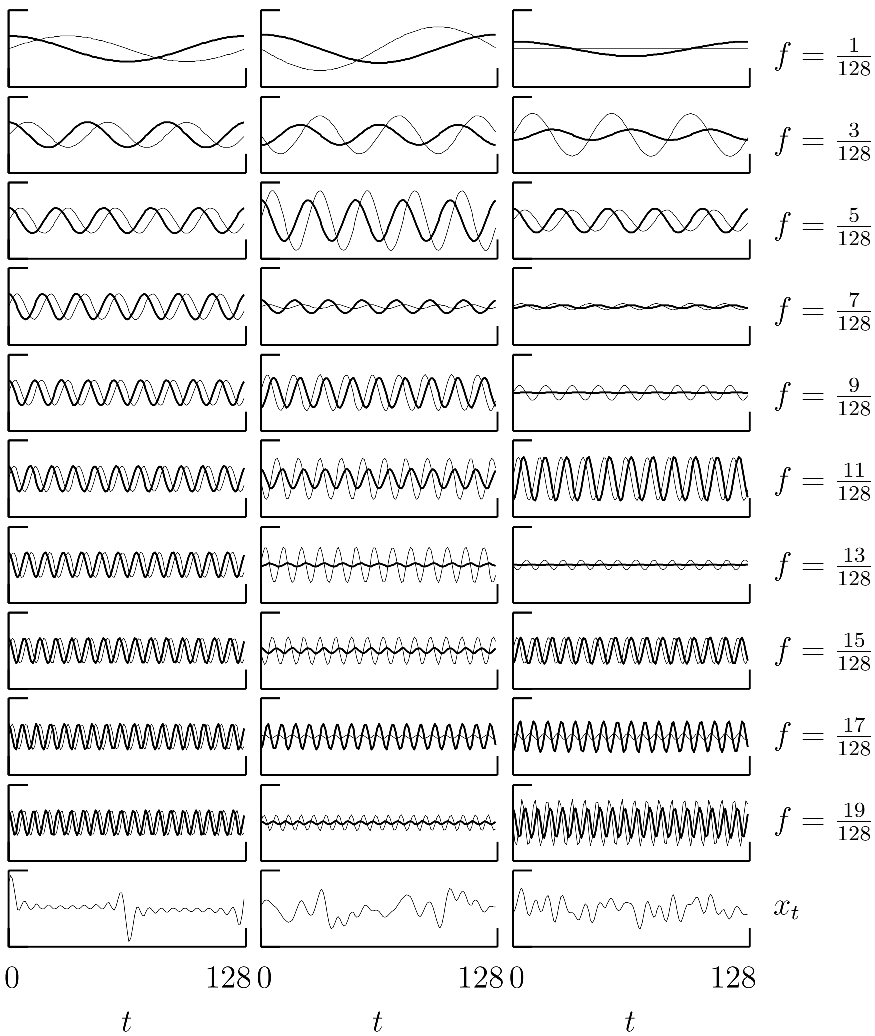


- now get series with 7 cycles over time span of 128
- in general plots of $\sin(2\pi\frac{k}{128}t)$ etc versus t have k cycles
- quantity $\frac{k}{128}$ is called frequency of sine or cosine (usually use variable f to denote frequency)
  - ◇ if k is small, sine time series is said to have low frequency
  - ◇ if k is large, sine time series is said to have high frequency
- amplitude (maximum range of variation) of $s_t = \sin(2\pi\frac{k}{128}t)$ is 1
- plots of $0.5\sin(2\pi\frac{7}{128}t)$ and $1.5\sin(2\pi\frac{7}{128}t)$



Sines and Cosines: III

- let's add together sines and cosines with different frequencies
  - ◇ first column uses sines and cosines of amplitude 1
  - ◇ second and third columns use random amplitudes



- conclusion: by summing up lots of sines and cosines with different amplitudes, can get artificial time series that resemble actual time series

Goal of Spectral Analysis

- given a time series x_t, figure out how to construct it using sines and cosines; i.e., to write

$$x_t = \sum_k a_k \sin(2\pi \tfrac{k}{128} t) + b_k \cos(2\pi \tfrac{k}{128} t)$$

- above called 'Fourier representation' for a time series (named after 19th Century French mathematician)
- allows us to reexpress time series in a standard way
- different time series will need different a_k's and b_k's
- can compare different time series by comparing the a_k's and b_k's

Gory Details: I

- Q: how do we figure out what a_k's and b_k's should be to form a particular time series?
- answer turns out to be surprisingly simple:

$$a_k = \frac{1}{64}\sum_{t=1}^{128} x_t \sin(2\pi \tfrac{k}{128} t) \text{ and } b_k = \frac{1}{64}\sum_{t=1}^{128} x_t \cos(2\pi \tfrac{k}{128} t)$$

- to see why this is reasonable, consider the following:
  - ◇ let $y_1, \ldots, y_{128}$ & $z_1, \ldots, z_{128}$ be collections of ordered values
  - ◇ let $\bar{y}$ and $\bar{z}$ be their sample means (i.e., $\bar{y} = \frac{1}{128}\sum_t y_t$ etc.)
  - ◇ let σ_y^2 and σ_z^2 be their sample variances: i.e.,

$$\sigma_y^2 = \tfrac{1}{128}\sum_{t=1}^{128}(y_t - \bar{y})^2$$

  - ◇ sample correlation coefficient:

$$\hat{\rho} = \frac{\frac{1}{128}\sum_t (y_t - \bar{y})(z_t - \bar{z})}{\sigma_y \sigma_z} = \frac{\sum_t y_t z_t}{128\sigma_y\sigma_z},$$

    where 2nd equality holds if $\bar{z} = 0$
  - ◇ measures strength of linear relationship between y_t's and z_t's ($-1 \leq \hat{\rho} \leq 1$)
  - ◇ relationship is strong if $\sum_t y_t z_t$ is large in magnitude
  - ◇ a_k is thus related to strength of linear relationship between $y_t = x_t$ and $z_t = \sin(2\pi\frac{k}{128}t)$, for which $\bar{z} = 0$

Gory Details: II

- to summarize how important frequency $\frac{k}{128}$ is in

$$x_t = \sum_k a_k \sin(2\pi \tfrac{k}{128} t) + b_k \cos(2\pi \tfrac{k}{128} t),$$

  let us form $S_k = \frac{1}{2}(a_k^2 + b_k^2)$

  - ◇ if frequency $\frac{k}{128}$ is important, then S_k should be large
  - ◇ if frequency $\frac{k}{128}$ is not important, then S_k should be small

- S_k over all frequencies $\frac{k}{128}$ is called the spectrum
- fundamental fact about the spectrum:

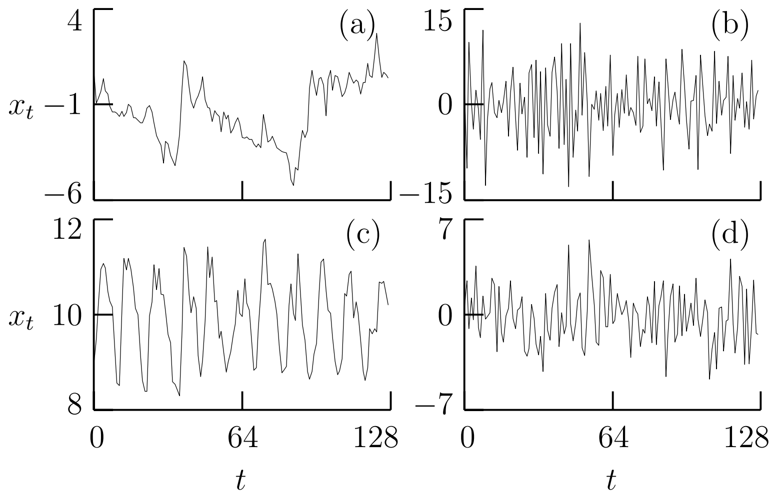
$$\sum_k S_k = \frac{1}{128} \sum_{t=1}^{128} (x_t - \bar{x})^2 = \sigma_x^2$$

  i.e., spectrum breaks sample variance of time series up into pieces, each of which is associated with a particular frequency

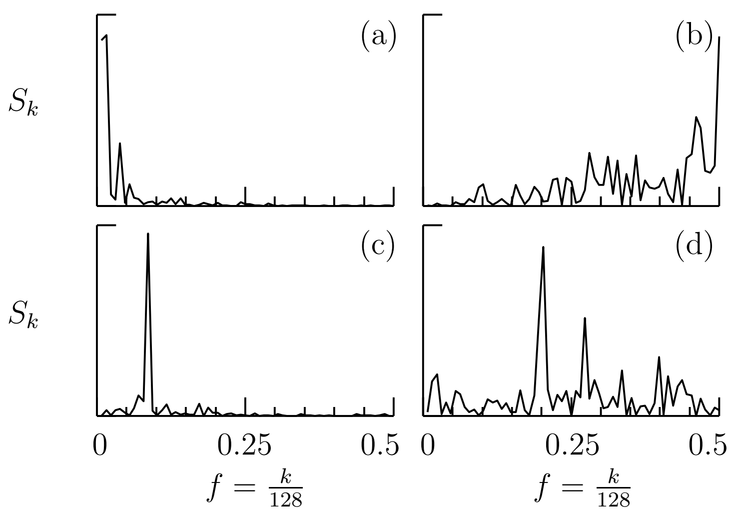
- spectral analysis is thus an analysis of variance technique, in which σ_x^2 is broken up across different frequencies

Examples of Spectral Analysis

- recall the four examples of time series



- here are the spectra for these four series



- uses include testing theories (e.g., wind data), exploratory data analysis (e.g., rainfall data), discriminating data (e.g., neonates), assessing predictability (e.g., atomic clocks)