Tutorial 1

Time Series & Data Analysis
NASSP – UCT

Due 24 February 2020 - 11am

You may use any programming language.

Submit both the code (can be as Jupyter notebook) and a pdf on cloudcape.saao.ac.za

Characterising data & smoothing

Question 1

For each of the following datasets,

- Sunspot data: `zuerich-monthly-sunspot-numbers.csv`

- Optical data of a Be X-ray binary: `F86_ogle234.dat`

- Confirmed cases of COVID-19: `novel-corona-virus-2019-dataset.zip`

- The spin frequency of X-ray pulsar GX 304-1: `gx304m1.fits`

(a) Plot the lightcurve or time series.

(b) Overplot the running mean.

(c) List the mean and variance.

(d) Make a histogram of the sampling.

(e) What is the mode of the sampling?

(f) Describe the time series in words.

Question 2

Use the `AirPassengers.csv` data for this dataset.

1. Detrend the AirPassengers data by using a moving average (rolling mean). Explain your choice of window, and plot the moving average over the data.

2. To explore seasonal variation in this dataset, you would remove the long term trend. Would you subtract it or divide it out? Substantiate your answer with some demonstrative plots.

Question 3

1. Code your own running mean function. Give some thought about how you will handle the start and end points of your time series.

2. Use your running mean function to detrend the `AirPassengers.csv` dataset.

Autocorrelation

You can use the `acf()` function in R or the `plot_acf()` function (among many others) in Python.

Question 4

Use the sunspot data (`zuerich-monthly-sunspot-numbers.csv`).

(a) What is the start, end and frequency sampling of the sunspot data?

(b) Extract the portion between April 1973 and December 1975. Resample it quarterly and plot it.

(c) For the full dataset, generate the ACF and plot it.

(d) Plot the time series against a copy that is lagged by 1 month.

(e) Plot the time series against a copy that is lagged by 11 years.

(f) What do these plots suggest about the stationarity of the data?

Question 5

Use the sunspot data (`zuerich-monthly-sunspot-numbers.csv`). It's clear that the cycle repeats on an 11 year timescale. Fold the sunspot data on this 11 year cycle, and plot

1. Sunspot number vs Phase

2. Average sunspot number vs Phase, in 30 phase bins