Procedures for CMI dataset

**Exploratory**

1. Import CMI data
2. Select `Year, Month, ­Day, Average Relative Humidity, Average Air Temperate (F)`
3. Replace missing temperatures with the mean of 7 days before and after
   1. Note: these days are outside the scope of our analysis later
4. Create time series plot for average temperature in CMI
   1. To understand the big picture of fluctuations
5. Calculate `Heat Index` = 83
   1. 85th percenticle of the 2 hottest months (July and August) between 1991-2021
6. Examine relative humidity and heat index time series plot
7. Create new variable `Heat Wave`
   1. Heat Wave = T when HI > 83 for 2+ consecutive days
   2. Find the instances of each heat wave and the dates from start to end
8. Plot the pulses, intensity, duration, and season length of heat waves
9. Find the initial parameters for nleqslv (For analysis section later on). These are the empirical parameters
   1. lambda = E[pulses]
   2. mu = 1/E[intensity]
   3. theta = 1/E[duration]
10. Create correlation pairs plots for the 4 variables

**Analysis**

Cleaning:

1. Repeat Exploratory steps 1 – 7
   1. only data cleaning portions
2. Select `date, heat\_index, heat\_wave\_85p`
3. Filter data between [05/01, 09/30] and [1989, 2020].
   1. Should end up with data of heat indexes between the months of the 30-year historical period

Modelling:

1. Calculate expected value and standard deviation of heat index
2. Create lag variable (the number of days ahead of day n). i.e.:
   1. X – random variable for heat index on day n
   2. Y – random variable for heat index on day (n + lag)
3. Find the means, standard deviation, covariance, and correlation
4. To solve for the theoretical parameters, apply Rodriguez-Iturbe derivations to a non-linear equation’s solver (using newton’s method) to solve for the three parameters at different lags
   1. Use Newton’s over Broyden’s as Jacobian matrix is not complicated to calculate and has faster convergence rates than Broyden’s method
5. Plot the different final theoretical values of the covariance and compare with empirical covariances

Project Abstract: **Heat Wave Trends in Illinois**

With global warming becoming a crucial issue worldwide, the goal of our project is to analyze current trends of heat waves in Illinois. To determine when heat waves are declared, we will use Illinois’ temperature and relative humidity to develop a heat index measurement. Next, statistical learning methods such as Rectangular Pulses Poisson Process model and Moving Block Bootstrap are implemented to investigate the variations of heat waves over several decades. The objective is to generate a model that describes the heat index of Illinois over our historical period.