**Renewable\_Power\_Statistical\_Model**

Statistical Models For U.S. Electricity Grid Renewable Production

Getting Data

1. Go to <https://nsrdb.nrel.gov/>
2. Open The NSRDB Viewer Tab
3. Use the 'Download Data' tab and select a specific location or area
4. Select all years, DHI, DNI, GHI, Cloud Type, and Wind Speed
   1. Ce sure to select leap days
5. Download data, a file containing the information will be emailed to the address provided
   1. This will download as a zip file and the data will have to be extracted
6. Store all extracted folders into one larger folder

1 - Combining Solar Data and Combining Solar Data Function

* Purpose
  + The data from the NSRDB is separated on different files for each year and location and it must be combined first. These two files work in tandem to combine all of the years at one location
  + The data is also cleaned and put into a tidy format
* Data Used
  + A folder that stores all of the raw data folders downloaded from the NSRDB
* Note

1. Open the Combining Solar Data Function and run the function into the environment
2. Open Combining Solar Data and change the file address to where the raw data is stored (lines 14 and 24)
3. Run the Combining Solar Data File
   1. This will save each locations combined data file to the working directory

2 - One Large File

* Purpose
  + Combines all data into one large file. Since the data is separated by location still
  + Calculates the daily averages of each field
* Data Used
  + Folder of all data created by the 'Combining Solar Data'

3 - The Wind Simulator

* Purpose
  + Fit the coefficients of the sinusoidal model that will predict the average wind speed for each day in the year
  + Identify the error distribution at each location
* Data Used
  + File of daily average weather statistics at each selected location ("Daily Averages per Location")
    - Created from the 'Combining Solar Data'
* Note

1. Preprocessing
   1. The data is normalized by subtracting wind speeds by the location's mean wind speed and dividing by the standard deviation
   2. Only 15% of all data is used to decreasing processing time
2. The model works by using thousands of combinations of sinusoidal coefficients and calculating the error with the actual value
3. The model with the least error is selected
4. Error distributions are then fitted. The cut off point between what errors follow a normal or gamma distribution is manually decided to roughly minimize the log-likelihood of both distributions
5. The 'local.num' needs to be increased by 1 after each location is fitted to avoid erasing any data

**-- KEY--** each location is fit one at a time due to long processing times and length of the code

4 - Simulated Wind Speed and Data Analysis

* Purpose
  + Use the models fitted with the 'The Wind Simulator' code to create simulated wind speed data for each location.
  + Then provide some insights on the wind speeds between each location. Primarily with correlation and cluster analysis
* Data Used
  + File of the best wind model parameters ("Wind Models - all locations")
    - Created from the 'The Wind Simulator' code
  + File of daily average weather statistics at each selected location ("Daily Averages per Location")
    - Created from the 'One Large File '
  + File of location codes and longitude and latitude of each location selected ("Locations with Lat and Long")
    - created from the 'One Large File'
* Note
  + The simulated wind speed is created by using the coefficients that were fitted by 'The Wind Simulator' and then inserted into the sinusoidal model with the same periods that predicts a wind speed for each day since the origin
  + Then a random amount of error is applied to each prediction. The random distributions are from 'The Wind Simulator' and are either normal or gamma some x% of the time that reflects the amount of error that were positively skewed.