# CSC110 Project Report:
# The Effect of Global Warming on Hurricane and Typhoon Occurrence

Mu "Samm" Du

Friday, December 14, 2020

## Introduction

The NASA Earth Observatory defines Global Warming as "the unusually rapid increase in Earth's average surface temperature over the past century" (Riebeek). As the global temperature has risen over the past decades, we are seeing more frequent occurrences of hurricanes and typhoons across the globe; these natural disasters have damaged properties, ruined communities, and destroyed people's lives (Miller). **Is the increase in hurricanes and typhoons related to global warming? If so, how are they related?**

I will analyze global temperature data procured by Berkeley Earth across land and ocean, as well as hurricane and typhoon occurrence data collected by the US National Oceanic and Atmospheric Administration limited to the Atlantic basin and the Pacific Ocean regions.

## Datasets

- Berkeley Earth — **The Berkeley Earth Land/Ocean Temperature Record**

  **Data format:** txt

  **Columns used:** Year, Month, Monthly Anomaly

  **Date range:** 1850-2020 (inclusive)

  **Source:** http://berkeleyearth.org/data-new/

- NOAA — **Atlantic hurricane database** (HURDAT2)

  **Data format:** csv in txt

  **Columns used:** ID, Date

  **Date range:** 1851-2019 (inclusive)

  **Source:** https://www.nhc.noaa.gov/data/hurdat/hurdat2-1851-2019-052520.txt

- NOAA — **Northeast and North Central Pacific hurricane database** (HURDAT2)

  **Data format:** csv in txt

  **Columns used:** ID, Date

  **Date range:** 1949-2019 (inclusive)

  **Source:** https://www.nhc.noaa.gov/data/hurdat/hurdat2-nepac-1949-2019-042320.txt

## Computational Overview

- I will complete the project in a Jupyter Notebook, so that I can present my code, descriptions, output, and interactive elements in a consistent format
- I will filter the datasets to only include relevant information to my analysis, such as the temperature values and timestamps in the global temperature data
- I will limit the time range such that the two datasets completely overlap in time
- I will aggregate the hurricane and typhoon data by counting the number of unique events that occur for every month. I will identify unique events by the ID column of each row, counting each unique ID as one event
- I will try to correlate global temperature with the number of hurricane and typhoon occurrence through a few regression functions, compare them, and pick out the one with the best fit without over-fitting;

  It is also important to acknowledge that while the temperature data is on a global scale, the hurricane and typhoon occurrence data is only limited to the Atlantic and Pacific regions, therefore a degree of error is involved with the correlation;

  I will use the **scikit-learn** library for the purpose of construction a regression model, since **scikit-learn** has a trove of regression functions built-in, and I will use **plotly** to present the data graphically, since it is a versatile plotting library that I have become familiar with in class
- To test the accuracy of my model, I will designate 8% of the data as the testing range, and compare the projection from my model with the data during that interval

## Obtaining Datasets and Executing Code

**Obtaining dataset files:**

- **Method 1:** You may clone the entire project and file structure from the following GitLab URL:

  `git clone https://gitlab.com/sammdu/csc110-final-project.git`
- **Method 2:** You may download the 3 dataset files (from the two original sources) from the following GitLab URL:

  `https://gitlab.com/sammdu/csc110-final-project/-/tree/master/data`

After downloading the 3 dataset files, you must place them under a folder named **data**, that is directly under the project root. You may reference the GitLab project from **Method 1**. You must name them the same way as is named in the GitLab project.

**View notebook online:**

https://nbviewer.jupyter.org/urls/gitlab.com/sammdu/csc110-final-project/-/raw/master/main.ipynb

**Executing code locally:**

The project code is written in a Jupyter Notebook file, `main.ipynb`. After installing all dependencies in the `requirements.txt` file, you can open a terminal in the project root and execute `jupyter notebook .` In the opened browser window, select **main.ipynb**. In the notebook file, simply execute every code block from top to bottom in order; alternatively, select "Kernel" ⇒ "Restart & Run All" to automatically run all code blocks in order. Make sure the correct Python 3 kernel is selected to avoid environment errors.

## Changes from Proposal

- No longer seek to predict future hurricane/typhoon events because correlation does not imply causation
- Using the NOAA data directly from NOAA website instead of Kaggle, for more recent data
- Isolate 8% of data for testing the reliability of the correlation

## Discussion

Despite significant variance in data, a weak direct correlation can be observed between the global temperature and hurricane/typhoon event count.

However, no regression model out of the two that were tested reliably predict the number of hurricane/typhoon events based on global temperature.

Between the two regression models, linear regression most accurately reflect the correlation between global temperature and the number of hurricane/typhoon events based both on $R^2$ scores and visual inspection.

For polynomial regression, a 3rd degree curve results in the smallest $R^2$ value on the Atlantic testing data, whereas a 2nd degree curve results in the smallest $R^2$ value on the Pacific testing data. However, these curves clearly exhibits signs of overfitting, and is worse at reflecting the correlation between global temperature and the number of hurricane/typhoon events compared to linear regression.

## References (MLA)

"Best Track Data (HURDAT2)." *NHC Data Archive*, NOAA, Apr. 2020, www.nhc.noaa.gov/data/#hurdat.

"Changes in Earth's Global Average Percentage Surface Temperature." *Berkeley Earth*, Accessed 11 Dec. 2020, berkeleyearth.org/data-new/.

"Documentation of Scikit-Learn." *Scikit-Learn*, sklearn.org/documentation.html.

Fincher, Jon. "Reading and Writing CSV Files in Python." *Real Python*, 21 Nov. 2020, realpython.com/python-csv/.

Jagannath@cr7. "Convert a Dictionary into a List in Python." *CodeSpeedy*, 11 Nov. 2019, www.codespeedy.com/convert-a-dictionary-into-a-list-in-python/.

Miller, Peter. "Weather Gone Wild." *National Geographic*, Sept. 2012, www.nationalgeographic.com/magazine/2012/09/extreme-weather-global-climate-change-effects/.

"Plotly Python Graphing Library." *Plotly*, plotly.com/python/.

Riebeek, Holli. "Global Warming." *Earth Observatory*, NASA, 3 June 2010, earthobservatory.nasa.gov/features/GlobalWarming.