Predicting severe storm events in the United States

Jenny Rhee

The goal of this capstone project is to predict the number of storm events aggregated at the state level. Building this model can potentially assist states/cities and the general population in preparation for severe storm events (e.g., infrastructure planning, stocking up on supplies, insurance, familiarizing with evacuation routes, etc.). People will likely have experience with these events, but the prevalence is expected to increase with climate change. It is important to begin understanding the rising trends because two or more severe storm events within a small window of time can be exceptionally more devastating (e.g., a severe flooding with a hurricane in the following days with no time to recover from the flood).

The National Weather Service provides storm data containing statistics on personal injuries and damage estimates from 1950 to present (https://catalog.data.gov/dataset/ncdc-storm-events-database). These events include, hurricanes, tornadoes, thunderstorms, hail, floods, drought conditions, lightning, high winds, snow, temperature extremes, etc. There are 51 columns including datetime, location, narrative, damage, etc. Due to the large number of types of storm events, initial analysis will use total count of storm events. Exploratory data analysis will begin with looking at the features available in the dataset and determine relevant features to start building the model. This includes visualizing any relationships, as well as the general trends over time per state and overall in the country.

Additional supporting data includes historical meteorological data to analyze any potential correlations. NOAA's National Centers for Environmental Information (ftp://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv/) has statewide, regional, and national temperature, precipitation, Standard Precipitation Index (SPI), and Palmer Drought Indices (PDSI, PHDI, PMDI, and ZNDX) from 1895 to present. Explanations for each index are:

- Palmer Drought Severity Index (PDSI) - monthly index that is generated indicating the **severity of a wet or dry spell**. This index generally ranges from -6 to +6, with negative values denoting dry spells and positive values indicating wet spells
    - > +/-4.0 = extreme wet spells/drought
    - +/-4.0 to +/-3.0 = severe wet spells/drought
    - +/-3.0 to +/-2.0 = moderate wet spells/drought
    - +/-2.0 to +/-1.0 = mild wet spells/drought
    - +/-1.0 to -/+0.5 = incipient wet spells/drought
    - -/+0.5 to 0 = normal

- Palmer “Z” Index (ZNDX) - “moisture anomaly index”; measure of the departure from normal of the moisture climate for that month
- Modified Palmer Drought Severity Index (PMDI) - the Palmer drought program calculates three intermediate parallel index values each month. Only one value is selected as the PDSI drought index for the month; based on probabilities (e.g., the probability that a drought is over is 100%, the probability that a wet spell is over is 100%, the probability is between 0% and 100%). **PMDI incorporates a weighted averaged of the wet and dry index terms, using the probability as the weighting factor.** PMDI and PDSI will have the same value during an established drought or wet spell (i.e., when the probability is 100%), but they will have different values during transition periods. This is the index that will be used for the model.

Because this dataset is relatively extensive, the initial goal is to predict the number of storm events per month. Depending on the outcome of the project, this may be extended to a higher resolution spatially and/or temporally. Jupyter Notebooks with code and explanations, an interactive map on my personal website, and a blog post will be the final products for this project.