



Github link : https://github.com/sstorey-nephila/w207_final_project/tree/master (https://github.com/sstorey-nephila/w207_final_project/tree/master)

Temperature Prediction Using Multivariate Time Series Analysis

The objective of this paper is to explain our methodology used to predict the temperature for a range of locations along the coast of mainland USA by applying machine learning with time series analysis. We begin by explaining the different features used today to identify temporal trends and patterns based on data from the "National Oceanic and Atmospheric Administration" (NOAA). We then move on to identify correlations across multiple time series that will be used to generate predicted values (weather forecast). We will explore the stack used to run our solution. Finally, we will implement a number of supervised methods and compare them to traditional techniques using "Autoregressive Integrated Moving Average Models" (ARIMA).

National Oceanic and Atmospheric Administration

"NOAA's National Centers for Environmental Information (NCEI) is responsible for preserving, monitoring, assessing, and providing public access to the Nation's treasure of climate and historical weather data and information" across the United States. it "hosts and provides access to one of the most significant archives on earth, with comprehensive oceanic, atmospheric, and geophysical data. From the depths of the ocean to the surface of the sun and from million-year-old ice core records to near-real-time satellite images, NCEI is the Nation's leading authority for environmental information."

Goals of this project

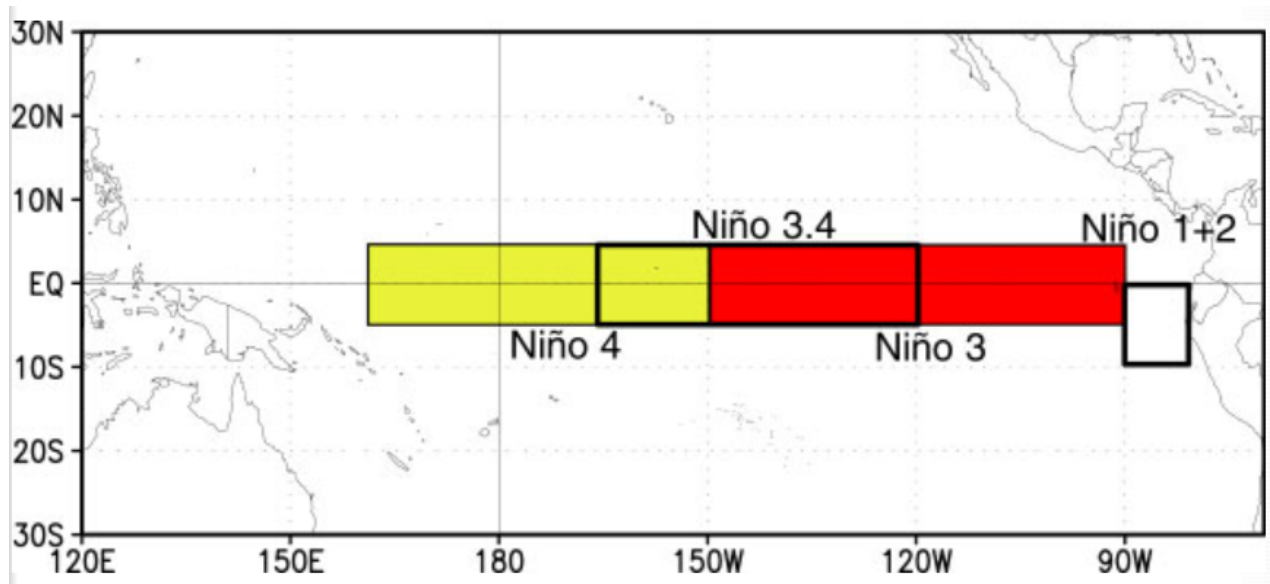
We will be attempting to

- Use teleconnections to help improve our temperature predictions across various locations in the USA
- Use feature engineering to extract key signals for our predictions
- Use a variety of machine learning techniques
- Automatically capture and log data about our experiments to aid repeatability

Background

Teleconnections in atmospheric science refer to climate anomalies that are related to each other at large distances (typically thousands of kilometers).

A prominent aspect of our weather and climate is its variability. This variability ranges over many time and space scales such as localized thunderstorms and tornadoes, to larger-scale storms, to droughts, to multi-year, multi-decade and even multi-century time scales.



This diagram shows the 4 key zones associated with the determination of if we are in an El Niño or La Nina phase

Some examples of this longer time-scale variability might include a series of abnormally mild or exceptionally severe winters, and even a mild winter followed by a severe winter. Such year-to-year variations in the weather patterns are often associated with changes in the wind, air pressure, storm tracks, and jet streams that encompass areas far larger than that of your particular region. At times, the year-to-year changes in weather patterns are linked to specific weather, temperature and rainfall patterns occurring throughout the world due to the naturally occurring phenomena known as El Niño.

Teleconnections & Weather Prediction

Teleconnection patterns are large-scale changes due to atmospheric waves that influence weather and temperatures as explained in the section above. In our project, we will be using sea surface temperature and 4 main teleconnection patterns from ENSO; AO (Arctic Oscillation), NAO (North Atlantic Oscillation), PNA (Pacific-North American Pattern), AAO (Antarctic Oscillation).

Data Sources

NOAA download tool <https://www.ncdc.noaa.gov/cdo-web/datatools> (<https://www.ncdc.noaa.gov/cdo-web/datatools>) <https://www.ncdc.noaa.gov/cdo-web/datasets> (<https://www.ncdc.noaa.gov/cdo-web/datasets>)

Data from 2000(by hour) <https://www.ncdc.noaa.gov/crn/qcdatasets.html> (<https://www.ncdc.noaa.gov/crn/qcdatasets.html>) <ftp://ftp.ncdc.noaa.gov/pub/data/uscrn/products/hourly02>

Daily from 1750(by day) <ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/readme.txt>
ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/ <ftp://ftp.cpc.ncep.noaa.gov/cwlinks/>
<ftp://ftp.ncdc.noaa.gov/pub/data/normals/1981-2010/source-datasets/>

Weather station index <ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-inventory.txt>
<https://www1.ncdc.noaa.gov/pub/data/ish/country-list.txt> (<https://www1.ncdc.noaa.gov/pub/data/ish/country-list.txt>)

Google BigQuery <https://cloud.google.com/bigquery/public-data/noaa-ghcn> (<https://cloud.google.com/bigquery/public-data/noaa-ghcn>)

Hourly data <ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-lite>

Monthly sst for el nino https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino12/
(https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino12/)

AMO <https://www.esrl.noaa.gov/psd/data/timeseries/AMO/> (<https://www.esrl.noaa.gov/psd/data/timeseries/AMO/>)

NINA34 - Anomaly from 1981-2010 https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino34/
(https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino34/)

References used

Information on teleconnections

http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/teleconnections.shtml
(http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/teleconnections.shtml)

Trick to use cosine/sine to encode seasonal components <https://www.kaggle.com/avanwyk/encoding-cyclical-features-for-deep-learning> (<https://www.kaggle.com/avanwyk/encoding-cyclical-features-for-deep-learning>)

Example using gated neural net https://github.com/Hvass-Labs/TensorFlow-Tutorials/blob/master/23_Time-Series-Prediction.ipynb (https://github.com/Hvass-Labs/TensorFlow-Tutorials/blob/master/23_Time-Series-Prediction.ipynb)

Feature engineering for time series <https://machinelearningmastery.com/basic-feature-engineering-time-series-data-python/> (<https://machinelearningmastery.com/basic-feature-engineering-time-series-data-python/>)

In [2]: `import nbconvert`