Accuracy Analysis of the National Weather Service (NWS) Weather Predictions

23 February 2018

### Objective

Use Cloud Computing Applications (CCA) to collect, analyze, and measure the accuracy of the weather forecasted by the [National Weather Service (NWS)](https://forecast-v3.weather.gov/documentation)[1] for different geographic locations in the United States.

### Background

Being able to predict the weather with a high amount of accuracy is an invaluable tool. According to a report published by [ForecastWatch](https://www.forecastwatch.com/static/High_Temperature_Accuracy_Study_12_Years.pdf)[2], the weather forecasts can predict tomorrow’s high temperature to within three degrees with 80% accuracy. They also found the forecasts reach the same accuracy 57% of the time, which is an improvement from a previous three-degree accuracy of 44%. This report lacks analysis of other aspects of weather, like precipitation, wind, and hazardous weather conditions.

In another report published in [The Huffington Post](https://www.huffingtonpost.com/richard-robbins/how-accurate-are-weather-_b_6558770.html)[3] examined the difference in precipitation forecast accuracy from the same day out to 10-day predictions at one location over the course of five months, with forecasted and actual precipitation data from the Weather Channel. In addition to only covering one location, this project also only covered one season, while it has been shown that prediction accuracy varies across different seasons (with accuracy being lower in November and December, for this particular region).

In this project, we will analyze the accuracy of the National Oceanic and Atmospheric Administration’s (NOAA) weather prediction abilities by locations.

### Data Pre-processing

The data for this analysis project come from the NOAA's [NWS](https://forecast-v3.weather.gov/documentation)[1] weather forecast website. NOAA offers existing datasets of historical weather observations, but not the predictions that were made for these dates. To get the predictions data, we would have to query the NWS Application Programming Interface (API)[1] every day for each location throughout the US. The data returned by this API will be in JSON format that contains the current weather and a week's worth of predictions. The sample python code in Appendix A prints the probability of precipitation for all the forecasted time periods in Urbana, Illinois.

In order to process and store the incoming JSON, we will utilize AWS instances. Also, each location could also be analyzed separately. A framework like Spark could be used to explore and analyze the data in parallel, without excessive waiting for results of queries.

### Research Questions

1. **Accuracy of precipitation predictions:** Of the days with a probability of at least 50% rain, what percent actually had rain? Also, the reverse for days with <50% rain probability.
2. **Accuracy and precision of temperature predictions:** Does the daily high temperature prediction tend to be higher or lower than what is actually observed? What is the variability of the difference between actual and observed temperature?
3. Which NOAA Offices regularly make better predictions than others? Which are the best and which are worst?

### References

1. NWS API Site: <https://forecast-v3.weather.gov/documentation>
2. Rose, Bruce and Floehr, Eric. 2017. Analysis of High Temperature Forecast Accuracy of Consumer Weather Forecasts from 2005-2016. Forecast Watch, LLC. <https://www.forecastwatch.com/static/High_Temperature_Accuracy_Study_12_Years.pdf>
3. Robbins, Richard. 2015. How accurate are weather forecasts? The Huffington Post. <https://www.huffingtonpost.com/richard-robbins/how-accurate-are-weather-_b_6558770.html>

### Appendix A: A Sample Python Code

import requests  
coords = { "Urbana, IL" : "40.1,-88.21"}  
url = "https://api.weather.gov/points/" + coords["Urbana, IL"]  
r = requests.get(url)  
url = r.json()["properties"]["forecastGridData"]  
r = requests.get(url)  
  
#print probability of precipitation for all the forcecasted time periods  
for period in r.json()["properties"]["probabilityOfPrecipitation"]["values"]:  
 print "{}: {}%".format(period["validTime"], period["value"])

Output:

2018-02-22T18:00:00+00:00/PT1H: 5%  
2018-02-22T19:00:00+00:00/PT4H: 6%  
2018-02-22T23:00:00+00:00/PT1H: 8%  
-------------Snipped--------------

### Appendix B: Software/CCAs

We plan to use the following software/CCAs for our analysis:

1. JSON
2. Python
3. Spark
4. AWS EC2 Instance
5. R

### Appendix C: Team Members

The following students will contribute to the group project:

|  |  |
| --- | --- |
| Name(First,Last) | NetID |
| Jordan Brooker | [brooker3@illinois.edu](mailto:brooker3@illinois.edu) |
| Justin Lavoie | [jlavoie2@illinois.edu](mailto:jlavoie2@illinois.edu) |
| Md Hasan | [mdhasan2@illinois.edu](mailto:mdhasan2@illinois.edu) (Contact person) |