Statistical Analysis of Weather Forecast Data: Wunderground Forecasts vs. Statistical Models

Yajie Chu, Sijia Huo, Yunxin Zhang, Yantong Zheng; Faculty Advisor: A.J. Hildebrand (University of Illinois at Urbana-Champaign)

UI Undergraduate Research Symposium, April 27, 2017

PREVIOUS WORK

- **Accuweather Long-Range Forecast Accuracy (Rosenberg, 2013) -** Forecasts collected from Accuweather over a 6-month period found to be worse than climatology past 10 days.
- **The Signal and the Noise (Silver, 2012)** - "Wet Bias": TV weather forecasters err toward overestimating the probability of rain.
- **How Valid Are T.V. Weather Forecasts?** (Freakonomics, 2008) - Kansas City TV stations found to provide less accurate forecasts than NOAA does.

DESCRIPTION OF DATA

Historical Data:

- **Data Source**: Utah State Climate Center from https://climate.usurf.usu.edu
- **Variables Collected**: Daily maximal and minimal temperatures for up to 100 years.
- **Locations**: From 4,625 locations in the U.S.
- **Size of Data**: 4.2 GB, over 100,000,000 data points altogether.

Wunderground Forecast Data:

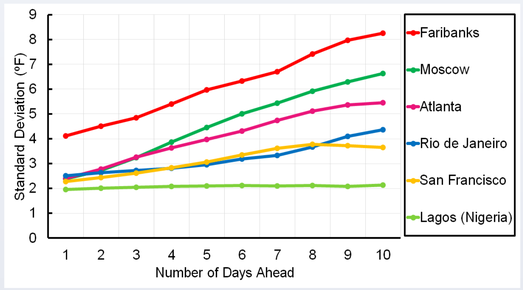
- **Data Source:** Collected each hour with API key from Wunderground https://wunderground.com using a Python script.
- **Variables Collected:** Hourly forecast of temperature and precipitation up to 240 hours from May 2016 to April 2017.
- **Locations:** For 33 selected international airports around the world.
- **Size of Data:** 2 GB, over 35,000,000 data points altogether.

METHODOLOGY

- **Persistence Model:** a model that assumes that the forecast temperature is the same as the current temperature.
- **Climatology Model**: a model that uses the historical average temperature for each day of the year as the temperature prediction for that day.
- **Combined Model:** an optimal combination of the persistence and climatology models.
- **Multiple Linear Regression (MLR) Model:** a linear regression model using temperatures of the previous 3 days as predictors.
- **Time Series Model:** an ARIMA model with parameters (3, 0, 3), which assumes the data to be stationary.

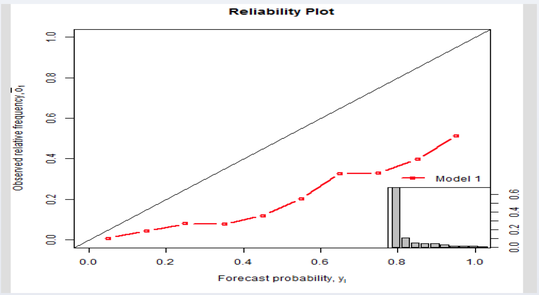
ACCURACY OF WUNDERGROUND FORECASTS

AVERAGE FORECAST ERROR BY LOCATION



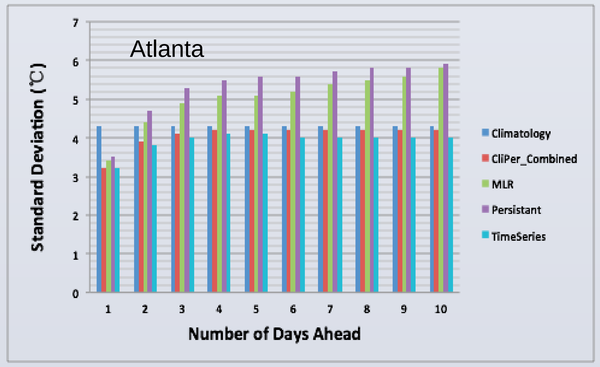
Average error for Lagos is consistently less than other locations; average errors for Chicago and Yakutsk indicate the most drastic increasing trends.

OBSERVED VS PREDICTED RAIN PROBABILITIES FOR CHAMPAIGN

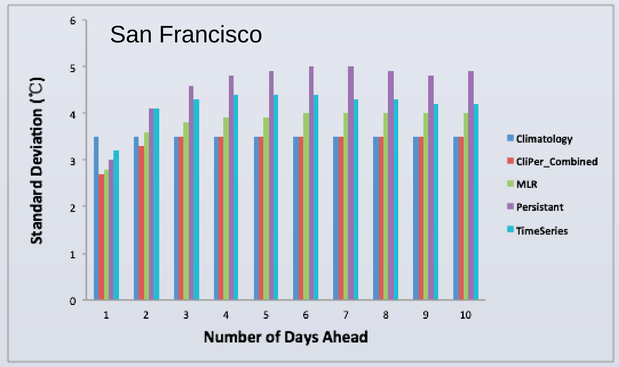


The graph shows the "wet bias": observed forecast probabilities are lower than predicted probabilities in the two-day forecast.

ACCURACY OF STATISTICAL MODELS

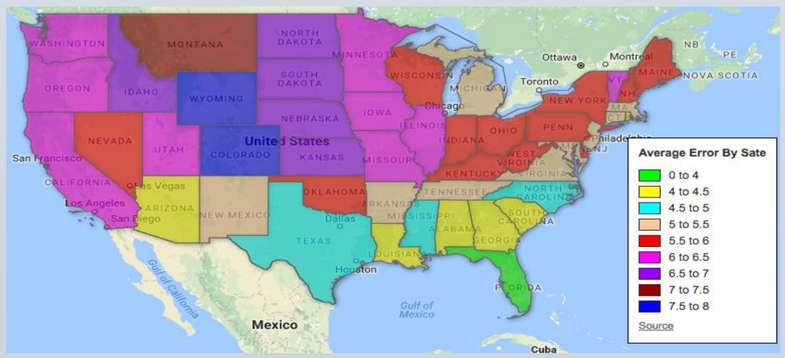


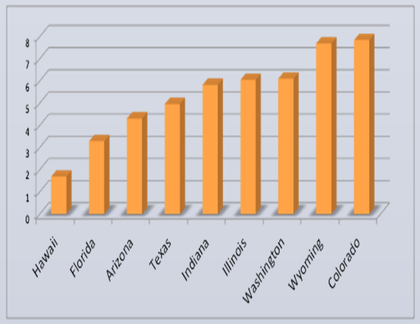
Combined and Time Series Models perform best.



Combined and Climatology Models perform best.

ACCURACY OF CLIMATOLOGY MODEL BY STATE





Weather is easiest to predict in Hawaii and Florida and hardest in Wyoming and Colorado.

SUMMARY OF OBSERVATIONS

- **Accuracy of statistical models**: An optimal linear combination of the Persistence and Climatology models (Murphy, 1992) performed better than the Persistence and Climatology models individually, but was not significantly different from the MLR and Time Series models (at the 1% level).
- **Accuracy of long-term forecasts**: In most locations, the average forecast errors increased at a linear rate from one day ahead to ten days ahead. In some locations (for example, Lagos and San Francisco), the ten-day forecasts were as accurate as the one-day forecasts.
- **Accuracy at different locations**: Within the U.S., all models perform better in southern states than in northern states. Among international cities, the accuracy was best at locations near the equator (for example, Lagos).

FUTURE WORK

- Some data violates the requirement of the Time Series to be stationary, so we need to solve the unstationary cases to obtain more complete predictions of this method.
- Scale and convert the Wunderground forecast in the form of Utah data to make the comparison between statistical models and online forecast more accurate.
- Diagnose why the average forecast errors of some locations decreases using data from 8-10 days ahead, for example, Beijing and Chicago.
- Extend the analysis to precipitation forecasts using the Brier Score.
- Extend the analysis to other forecast sites, for example, AccuWeather.com.

REFERENCES

- **J.D. Eggleston and S. Dubner, How Valid are T.V. Weather Forecasts**, Freakonomics.com, 12 April 2008
- **B. Hanson, Internet Weather Forecast Accuracy**, OmniNerd.com, 07 February 2007
- **A. Murphy, Climatology, Persistence, and Their Linear Combination as Standards of Reference in Skills Scores**, Weather and Forecasting 7: 692-698, 13 July 1992
- **J. Rosenberg, Accuweather long-range forecast accuracy questionable,** Josh-Rosenberg.com, 13 June 2013
- **H. Stern, The accuracy of weather forecasts for Melbourne, Australia,** Meteorol. Appl. 15: 65-71, 24 January 2008
- **N. Silver, The Signal and the Noise,** 27 September 2012