# Evaluation of weather forecast accuracy

By Evan Kennedy
Mentored by William deRosset

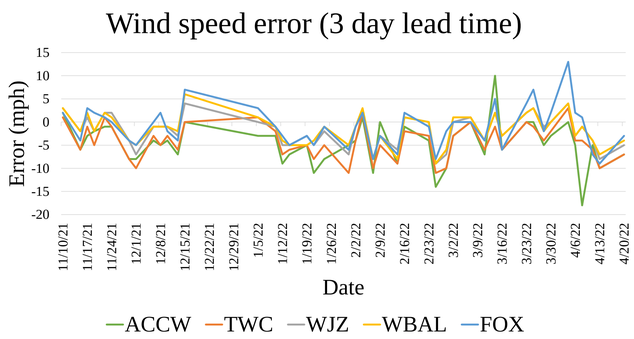
## Introduction

Weather forecasts are an integral part of modern life, aiding in the planning of events, public safety decisions, and economic planning. Because of the crucial role they play in the modern society, the accuracy of weather forecasts are incredibly important. Relying upon inaccurate weather forecasts can result in unnecessary casualties and economic losses (Cook & Gruebacher, 2008). The aim of this project was to evaluate and compare the accuracy between different weather forecasting services to determine the reliability of each service in relation to its competitors. Similar research has been performed by the company ForecastWatch, which found that modern forecasts are highly accurate (Rose & Floehr, 2017), results this study was expected to replicate in findings of low forecasting error and low variation between forecasters.

## Materials & Methods

The forecasting services analyzed for this study were AccuWeather, The Weather Channel, WJZ Baltimore, WBALTV, and FOX 45. The accuracy of forecasts from each source were analyzed for the Baltimore area from October 2021 to March 2022. Forecasts from each source were documented daily during the testing period and were compared for accuracy to the observed weather readings from the ASOS weather station at BWI airport. Parameters analyzed from each forecaster were high temperature, low temperature, probability of precipitation, and wind speed, with an additional parameter for the time scale of the forecast being analyzed, including forecasts of lead times one through five days for all forecasting sources and weather events.

Forecast error was calculated for high temperature, low temperature, and wind speed by subtracting the forecasted value from the observed value for the appropriate corresponding day. Forecast error for probability of precipitation was calculated for groups based on percentage of forecast prediction, with four groups being formed for analysis on intervals of 25% from 0% to 100%, and then, the frequency of days with measurable precipitation (> 0.01 in.) for the forecasted days within a group were subtracted from the average percentage forecast for the group to produce an error value for the percentage group. An example of error values can be seen in Graph 1 below for wind speed at a lead time of three days.




Graph 1 (left): Absolute error of wind speed at lead time of three days. Note that while the graph encompasses the period of the study, data is missing between 12/15/21 and 1/5/22.

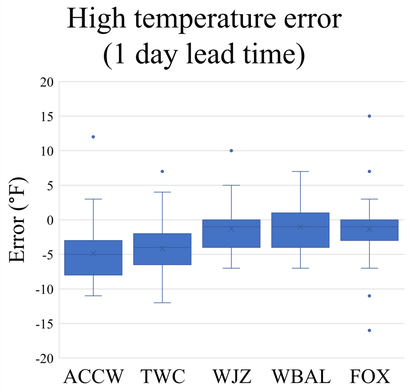
## Materials and Methods (cont.)

Error analysis for a forecast group (all forecasts from the same source, for the same weather event, for the same lead time) was performed by calculating two values: mean error and root mean squared error. Mean error is the average of the error values accounting for positive and negative error values, indicating a bias in forecasting towards overshooting or undershooting the actual weather occurrence, while root mean squared error is the average of absolute error values, indicating overall accuracy for a forecast group. To determine if differences between sources in forecast error indicated a significant difference in forecast accuracy, ANOVA statistical tests were run between the error of the forecasting groups to determine if there was a statistically significant difference in error for the given weather event and lead time being analyzed, with a statistically significant result indicating that one or more forecasters were notably more or less accurate for the given forecast parameters.

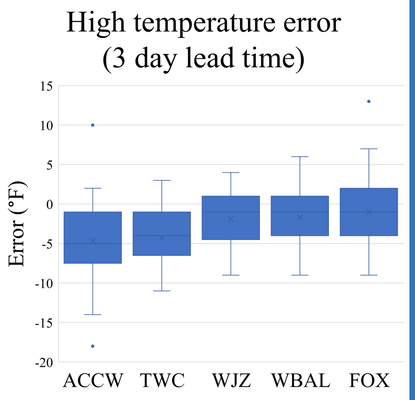
Materials used for this study were the applications Excel and Minitab for data processing and statistical analysis, as well as access to the websites of the specified weather forecasting services for data collection.

## Results

ANOVA tests that yielded a significant $p < .05$ were those performed for high temperature error, with lead times of one [$F(4, 225) = 10.30$, $p < .001$], two [$F(4, 225) = 7.95$, $p < .001$], three [$F(4, 225) = 7.65$, $p < .001$], and four [$F(4, 225) = 5.10$, $p = .001$] days, indicating a significant difference in accuracy between forecasting services for those parameters. The distribution of accuracies for two parameter combinations, high temperature forecasts at lead times of one and three days, that yielded significant difference between forecasters can be seen in Graph 2 and 3.




Graph 2 (above): Box plot displaying distribution of forecast error for all five studied forecasters for high temperature forecasts of lead time one day.




Graph 3 (above): Box plot displaying distribution of forecast error for all five studied forecasters for high temperature forecasts of lead time three days.

## Results (cont.)

ANOVA tests run for all other weather event/lead time parameter combinations yielded $p > .05$, indicating no significant difference in accuracy between forecasting services for those parameters.

Post-hoc Tukey tests were performed on the high temperature/lead time parameter pairs that exhibited significant differences between weather forecasting services. These post-hoc tests revealed a consistent pattern of significance, with both AccuWeather and The Weather Channel being significantly different from WJZ Baltimore, WBALTV, and FOX 45 while not being statistically different from each other, for all parameter pairs exhibiting significant differences between forecasters.

## Conclusions

The purpose of this study was to evaluate the accuracy of different weather forecasting services in order to determine the reliability of those forecasters. From ANOVA tests, it was found that there was no significant difference in accuracy between the analyzed weather forecasting services for all analyzed weather events excluding high temperature, for which a significant difference in accuracy between forecasters was exhibited, for all lead times less than five days. For high temperature forecasts, the larger scale weather forecasting analyzed for the study (AccuWeather and The Weather Channel) were significantly less accurate than the local stations analyzed for the study (WJZ Baltimore, WBALTV and FOX 45), averaging error about 2 °F greater in magnitude than their counterparts for high temperature forecasts. While only a significant pattern for high temperature, this consistent trend towards comparatively greater accuracy of local forecasters is worth verifying and potentially explaining, especially given that the trend contradicts what would be expected following the assumption of weather forecast accuracy, particularly for high temperature forecasts, presented by ForecastWatch. The success of local forecasters could be attributable to a multitude of factors, and extensions of this study to encompass more forecasters servicing a larger area over a longer time frame could serve to determine the methodologies and/or practices that may potentially contribute to superior local forecasting accuracy, knowledge that could valuably inform the practices of forecasting professionals and the consumption choices of forecast consumers.

## References

Cook, R. K., & Gruenbacher, B. (2008, October 20). Assessment of methodologies to forecast wind gust speed. *National Weather Service*. https://www.weather.gov/ict/windgust

Rose, B., & Floehr, E. (2017). Analysis of high temperature forecast accuracy of consumer weather forecasts from 2005-2016. *ForecastWatch*. https://www.forecastwatch.com/wp-content/uploads/High_Temperature_Accuracy_Study_12_Years.pdf