

Improvement of Daily Temperature Forecasts from a Prediction Market

How ForecastEx prices for the daily high compare with 20 forecasting systems at 24 US locations, 11 February to 1 October 2026.

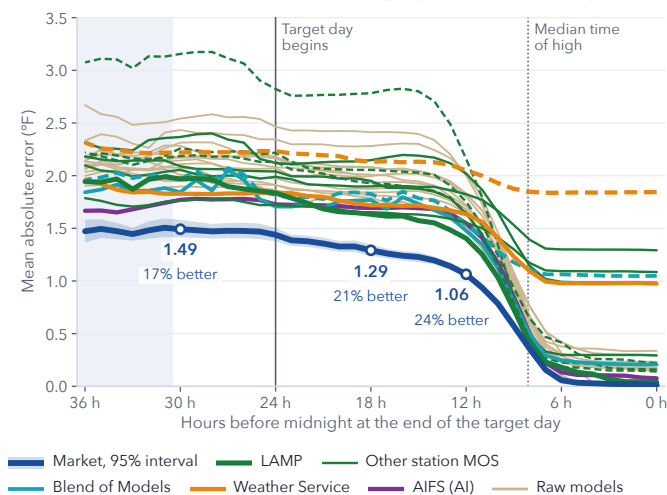
Patrick T. Brown · Trend Analytics, LLC and Interactive Brokers · Johns Hopkins University · The Breakthrough Institute

OVERVIEW

- Prediction markets have a documented history of distilling dispersed information into a well-calibrated price, the wisdom of the crowd.
- They are thought to add the most value on subjective questions such as elections, where information is scattered and human biases matter.
- It is far less clear that they should improve on weather forecasts, which rest on physical and statistical models refined over decades and supervised by meteorologists.
- Against 20 established systems at 24 US locations over 5,405 location-days, the ForecastEx market has the lowest error at all meaningful lead times before the high was observed.

ERROR THROUGH THE FORECAST WINDOW

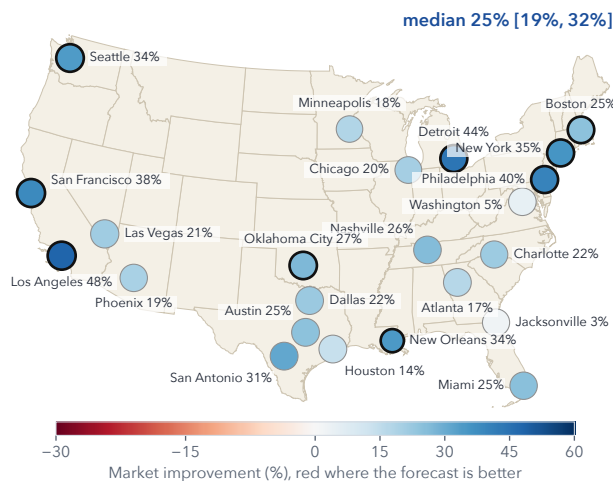
Mean absolute error of each forecast of the daily high, all 24 locations together



Blue, the market with its 95% interval; dashed, a system scored a second way. Under the market's error, its improvement over the next-best alternative. After the high the alternatives are held at the temperature reached and converge; shaded leads cover only some days.

LOCATION BY LOCATION, 30 HOURS AHEAD

How much lower the market's error is than the National Weather Service's

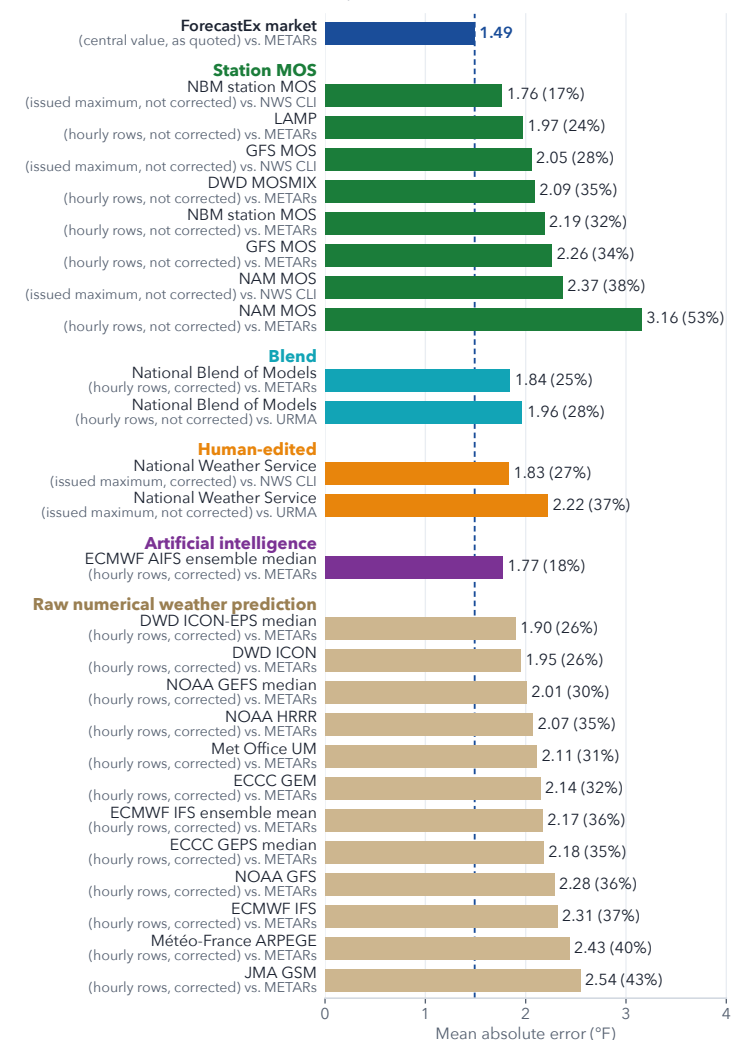


Each circle is a location, labelled with the market's improvement, sized by days scored and bold where the difference holds after a multiple-comparison adjustment. The forecast is scored against the Weather Service's own climate report. Ahead at all 23 locations, 9 in bold (Buckley has no climate report).

- From the evening before to six hours before the high, its errors are **17% to 24%** below the next-best alternative's, and **27%** below the National Weather Service forecast 30 hours ahead.
- Its margin over the 20 systems is 32% against forecasts of any age and 32% against freshly issued ones, so the edge is accuracy at the moment of forecasting.
- Beating the statistical products built to correct the models, and the human-edited forecast, suggests the market combines information more flexibly than conventional forecasting.

EVERY SYSTEM, 30 HOURS AHEAD

Error (°F) and, in brackets, the market's improvement over each of the 25 series



$$\text{Improvement} = 1 - \frac{\text{MAE (ForecastEx vs. METARs)}}{\text{MAE (system vs. its own standard)}}$$

Each system is held to the standard it is built to forecast; for the National Weather Service the denominator is MAE (NWS NDFD vs. NWS CLI), its climate report, and its second scoring uses URMA. Each label ends with that standard. Issued maximum, the system's own daily high; hourly rows, the highest of its hourly forecasts; corrected, each location's recent bias removed. METARs, the weather reports the market settles on; NWS CLI, the Weather Service's climate report; URMA, NOAA's gridded analysis. All 25 differences are significant.