ForecastWatch

DATA ANALYSIS REPORT • 12-YEAR FORECAST ACCURACY

Analysis of High Temperature Forecast Accuracy of Consumer Weather Forecasts

2005 – 2016

Authored by

Bruce Rose & Eric Floehr

Data provided by ForecastWatch, a service of Intellovations, LLC

Intellovations, LLC

info@forecastwatch.com • www.forecastwatch.com

6724 Perimeter Loop Road, #275 • Dublin, OH 43017

614-923-2050 • 855-609-9609

About ForecastWatch

ForecastWatch, a service of Intellovations, LLC, has been the world's premier weather forecast monitoring and analytics company since 2003. Our passion for data drives us every day. We collect weather forecast data from thousands of locations throughout the U.S. and around the world. This information is added to an ever-growing and unparalleled historical database of more than 800 million weather forecasts gathered from a variety of resources.

We use this vast collection of data to evaluate and compare weather forecast providers, improve decision-making by governments and business entities impacted by weather, improve weather forecasting by meteorologists around the world and educate customers with unbiased reporting.

We strive to improve and expand our offerings to meet the needs of our current and future clients, finding ways to partner with them to help them evaluate their own deliverables, keep their customers safe or help make business-critical decisions by analyzing weather forecasts to positively impact revenue, operating costs and risk mitigation costs.

Meteorologists, utilities and energy companies depend on ForecastWatch's accurate data and analysis. Agriculture, futures traders and other companies whose business depends on being right about the weather put their trust in us to help them achieve success. Even consumers benefit from our ForecastAdvisor product. Our data meets the highest standard of scientific inquiry and has been used in several peer-reviewed studies.

Executive Summary

This report analyzes forecast accuracy and trends over a twelve-year span of time from ten different forecast providers. The goal is to answer the question, “How accurate are weather forecasts?” Specifically, this report analyzes high temperature forecasts between one and ten days in advance, and looks at error, bias, and trends in the accuracy of those forecasts. Graphs and discussion are included.

ForecastWatch data show that high temperature forecasts are generally extremely accurate, and continue to measurably improve.

Key Findings

1 **One-day-out forecasts are extremely accurate.** Today's forecasts average under 3°F error.

2 **One-day-out forecasts have improved substantially** over the past twelve years, with error declining by 33% over the analyzed time period.

3 **Five-day-out forecasts gained the most accuracy.** Today they nearly match the accuracy of one-day-out forecasts at the start of the study in 2005.

4 **Nine-day-out forecasts** have only recently become slightly better than long-term climatological average data.

5 **Forecasts generally predict warmer than actual temperatures (positive bias),** but this bias is declining.

At a Glance

33%	40%	~3°F	200M
Reduction in 1-day-out error over 12 years	More perfect 1-day-out forecasts than in 2005	Avg. 1-day-out error in 2016	Forecast verifications analyzed

Analysis & Methods

Approach

High temperature forecasts from one- to ten-days-out were compared with observed high temperatures. Pairs of forecast and observed data were assessed via root mean square error (RMSE), which is a standard metric of forecast accuracy. Pairs were also categorized by absolute error as “Perfect” (<1°F error); “Good” (≤3°F error); and “Bust” (≥10°F error). Analyses were performed to assess accuracy over time, bias, and differences in accuracy between near- and long-term forecasts.

Data Set

Analysis was based on forecast data provided by ForecastWatch. The forecast data contained nearly 200 million high temperature forecast verifications for nearly 800 locations in the U.S. over a twelve-year period. Data included forecast high temperatures for up to ten days into the future, depending on the provider. Actual temperature observations were collected from the National Climate Data Center (NCDC) from official observation records of the ASOS/AWOS observation network.

Data Sourcing

Daily forecasts were collected using web crawler software to inspect and “scrape” data from public websites of forecast providers, or collected from public or private APIs made available by those providers. ForecastWatch monitored ten popular providers: AccuWeather, Foreca, Intellicast, MeteoGroup, CustomWeather, The Weather Channel, Weather Underground, the National Weather Service, an anonymous private weather forecast provider and a private feed from Global Weather Corporation. Additionally, a baseline reference forecast was created from 1971-2000 climate averages. The collection process was run daily, starting at 22:00 Coordinated Universal Time (5 p.m. Eastern Standard Time or 6 p.m. Eastern Daylight Time). The process generally took 30 minutes to complete each day.

Actual temperature observations were collected by ForecastWatch from the National Climate Data Center (NCDC). NCDC makes official observations available each day from a network of high-quality observation stations that are part of the Automated Surface Observing System (ASOS) and Automated Weather Observing System (AWOS) network. These observations typically take place at major airports, weather offices, and other landmarks (for example, New York's Central Park or Astoria, Oregon).

Data Completeness

Not all forecast lengths and dates are available for all providers through the entire twelve-year time span, for a variety of reasons: the number of forecast providers has increased with time; not all providers make forecasts at all time lengths; and technical issues, such as changes in data availability policies, may temporarily limit access to forecast data.

ForecastWatch

Forecast & Observation Pairs by Year

Table 1 below shows the number of forecast and observation pairs used for each year and forecast period in the study. Available forecasts have increased from about 1.5 million forecast and observation pairs in 2005 to nearly 2.3 million forecasts in 2017. Size change in this data set occurred when a provider was added or removed to the analysis, in part (new locations or extended days-out forecasts) or in whole (CustomWeather blocked ForecastWatch from collecting in 2015).

Year	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Total
2005	1,503,396	1,502,944	1,496,516	1,458,229	1,191,453	1,189,689	857,495	847,411	841,944	**10,889,077**
2006	1,775,365	1,776,071	1,771,851	1,707,222	1,524,353	1,524,498	1,013,717	1,005,889	1,005,030	**13,103,996**
2007	1,750,395	1,750,242	1,745,262	1,721,480	1,548,139	1,529,457	1,030,512	1,026,001	1,000,239	**13,101,727**
2008	1,964,344	1,962,679	1,954,923	1,931,053	1,584,443	1,489,761	1,260,510	1,258,532	1,034,556	**14,440,801**
2009	2,118,941	2,118,504	2,111,635	2,086,654	1,604,515	1,454,162	1,342,242	1,342,293	1,117,359	**15,296,305**
2010	2,124,523	2,123,867	2,121,371	2,090,830	1,608,953	1,516,128	1,344,408	1,344,365	1,343,762	**15,618,207**
2011	2,270,628	2,270,083	2,267,848	2,231,388	1,931,928	1,860,483	1,414,124	1,413,262	1,412,584	**17,072,328**
2012	2,558,043	2,556,493	2,553,978	2,513,172	2,308,577	2,195,309	1,721,041	1,720,668	1,719,802	**19,847,083**
2013	2,551,806	2,551,354	2,550,178	2,513,392	2,301,996	2,251,864	1,799,597	1,799,772	1,799,292	**20,119,251**
2014	2,599,516	2,599,288	2,598,933	2,562,201	2,346,244	2,292,616	1,828,021	1,830,099	1,830,077	**20,486,995**
2015	2,391,370	2,392,480	2,384,241	2,330,542	2,141,903	2,075,827	1,621,671	1,622,363	1,612,728	**18,573,125**
2016	2,324,966	2,324,617	2,323,583	2,233,885	2,072,189	1,920,734	1,554,515	1,554,519	1,554,579	**17,863,587**

Table 1 • Number of forecast and observation pairs in the ForecastWatch data set by year and day of forecast

Date Definitions

The high temperature forecast data set contains forecasted values for a specified number of days into the future, from the current day to nine or more days into the future. A one-day-out high temperature forecast is the forecast for the next day. For example, for a forecast collected on January 1, 2016, the one-day-out high temperature forecast would be the forecast for January 2, 2016.

Calculation of Error

Pairs of forecasts and observations were selected in one-month batches. Observations and forecasts were compared using root mean square error (RMSE) and mean error (ME). RMSE provides a single measure of overall forecast accuracy, and is a commonly cited statistic when assessing forecast accuracy. RMSE cannot detect systemic high or low bias in forecasts. ME is used to assess such bias.

Calculation of Error (continued)

To calculate RMSE, the arithmetic difference of a high temperature forecast-observation pair is squared, then summed, and divided by the number of events. The square root of the sum is the average squared error. Lower RMSE means more accurate forecasts; higher RMSE means less accurate forecasts. A perfect set of forecasts would have RMSE equal to zero.

Mean Error is similar to RMSE but does not square the forecast-observation difference. The arithmetic difference is summed, then divided by the number of events. Lower ME (either positive or negative) means less biased forecasts; higher ME means more bias in the forecasts. A perfectly unbiased set of forecasts would have ME equal to zero.

Squaring error has the effect of penalizing forecasts with large variance of error. This effect is desirable since it more harshly penalizes less reliable forecasts. However, a set of forecasts may be highly inaccurate (high RMSE), yet not be biased (zero ME). Such a set is not biased, but is nonetheless poor in accuracy. In contrast, another set of forecasts may be highly accurate (low RMSE), yet have consistent bias (high ME). If the nature of the bias is known, compensations may be made.

Forecast Categorization

The study categorizes forecasts into three categories and calculates the percentage of forecasts that fall within a particular category:

PERFECT	GOOD	BUST
< 1°F	**≤ 3°F**	**≥ 10°F**
Forecast error less than 1°F	Forecast within ±3°F of observation	Forecast in error by at least ±10°F

Introduction

Weather forecasters and meteorologists receive a lot of (occasionally good-natured) grief about the accuracy of weather forecasts. “Six inches of partly cloudy!”; “Must be great to work in a profession where you only need to be right half the time!”; “I could predict the weather better!” Despite the jokes, people rely on weather forecasts and weather information extensively. While there may not be social consensus on the accuracy of forecasts, forecast data are more available than ever before. Is the conventional wisdom correct? Are forecasts no better than long-term averages?

Introduction (continued)

This report presents an in-depth analysis of the accuracy of weather forecasts, specifically focusing on high temperature predictions. The goal of the analysis was to assess the accuracy of forecasts, and to determine the degree to which forecasts are improving with time. The data overwhelmingly confirm what experts already know: weather forecasts are highly accurate, and are improving dramatically.

Technology is a primary driver in the improvement in forecast accuracy. Not only does technology make forecasts more accessible, it also makes forecasts more powerful and precise. New satellite, radar, and ground-based sensors have made weather observations better than ever, with unprecedented density and resolution around the globe. The higher quality data combined with advanced computational platforms have enabled the proliferation of new and superior computer models to predict weather further into the future. The availability and accuracy of models and other tools have enabled human forecasters to improve both their knowledge and skill, resulting in superior forecasts.

This report was generated by ForecastWatch. We've been measuring the accuracy of consumer weather forecasts since 2005. Each day, we check the accuracy of the most popular weather forecasters at predicting high and low temperatures. We collect data at thousands of locations across the U.S. and the world. We compare reports of precipitation, cloud cover, wind, and the chance of precipitation at the same locations. Over the past twelve years, we have amassed a large database of forecast-observation pairs.

Results

All forecasts, from short- to long-term, are improving

Forecast error, percentage of busted forecasts, and percentage of perfect forecasts all improved over the study period, for all forecast lengths from one- to seven-days out. High temperature forecast error improvement ranged from ~1°F for one-day-out to nearly 2°F for five-day-out forecasts over the twelve-year study period, as shown in Figure 1.

- In 2005, one-in-thirty (3%) one-day-out and one-in-five seven-day-out forecasts were “busts” (>10°F error). By 2016, busts were reduced to one-in-70 (1.5%) and one-in-eight (13.3%), respectively (Figure 2).
- In 2005, a six-day-out forecast would have had a one-in-five chance of having at least a ten-degree error. Twelve years later, the likelihood has decreased to one-in-ten (Figure 2).
- The percentage of perfect (<1°F error) one-day-out forecasts improved from 11.3% to 15.7% over twelve years, which is a 40 percent improvement in the number of perfect forecasts (Figure 3).

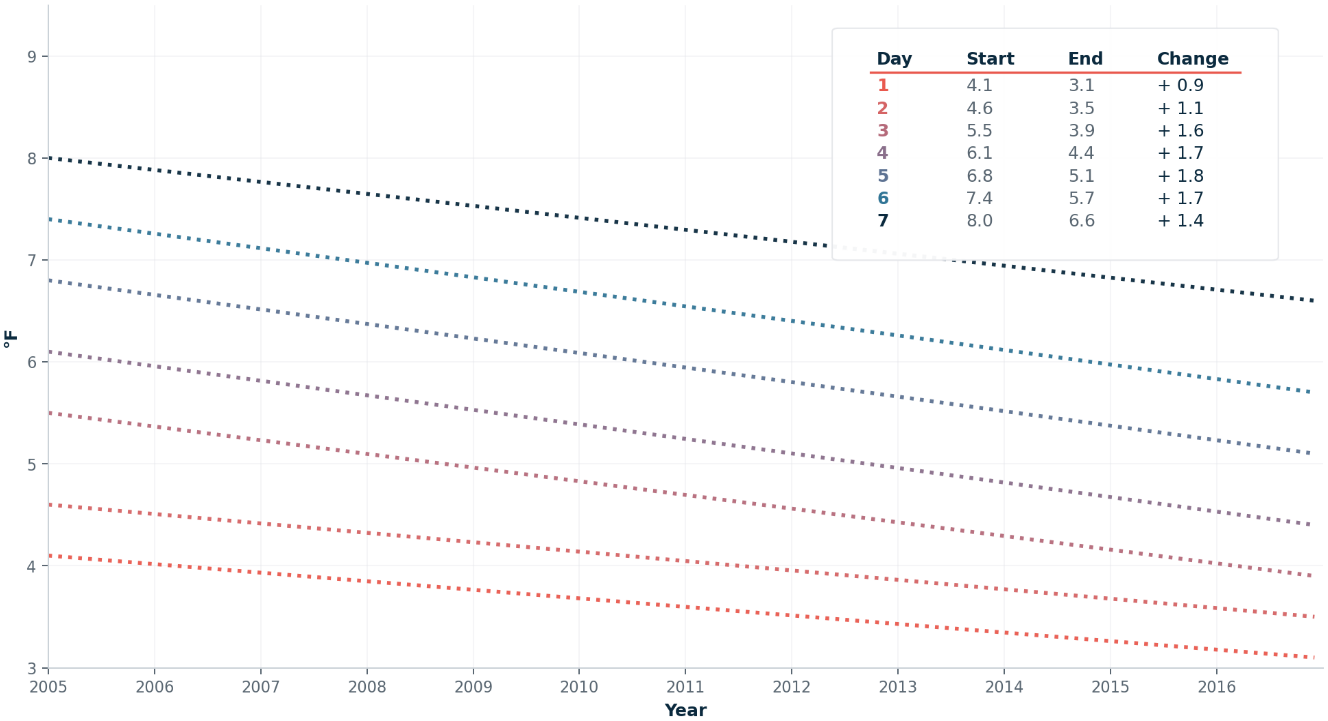

Figure 1 • Day 1–7 RMSE Trendlines by Year, 2005–2016

Results

Busts down, perfect forecasts up

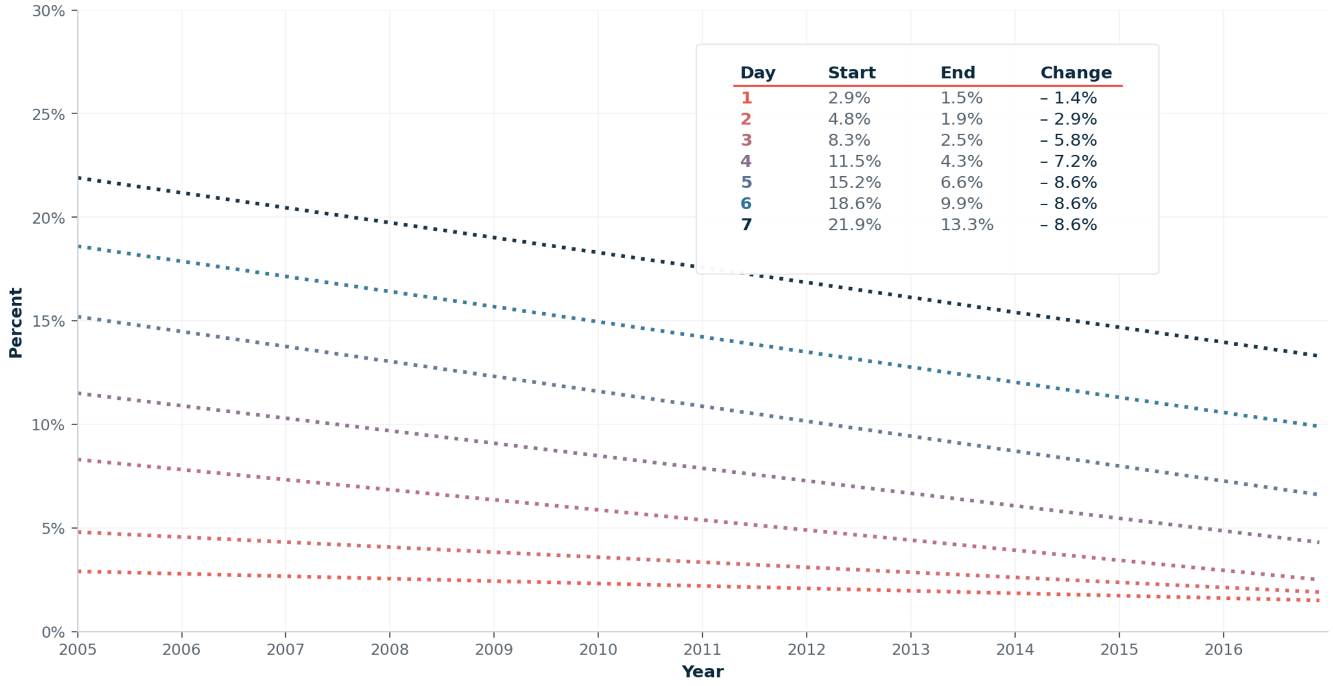

Figure 2 • Day 1–7 Busted Forecast Percentages by Year, 2005–2016

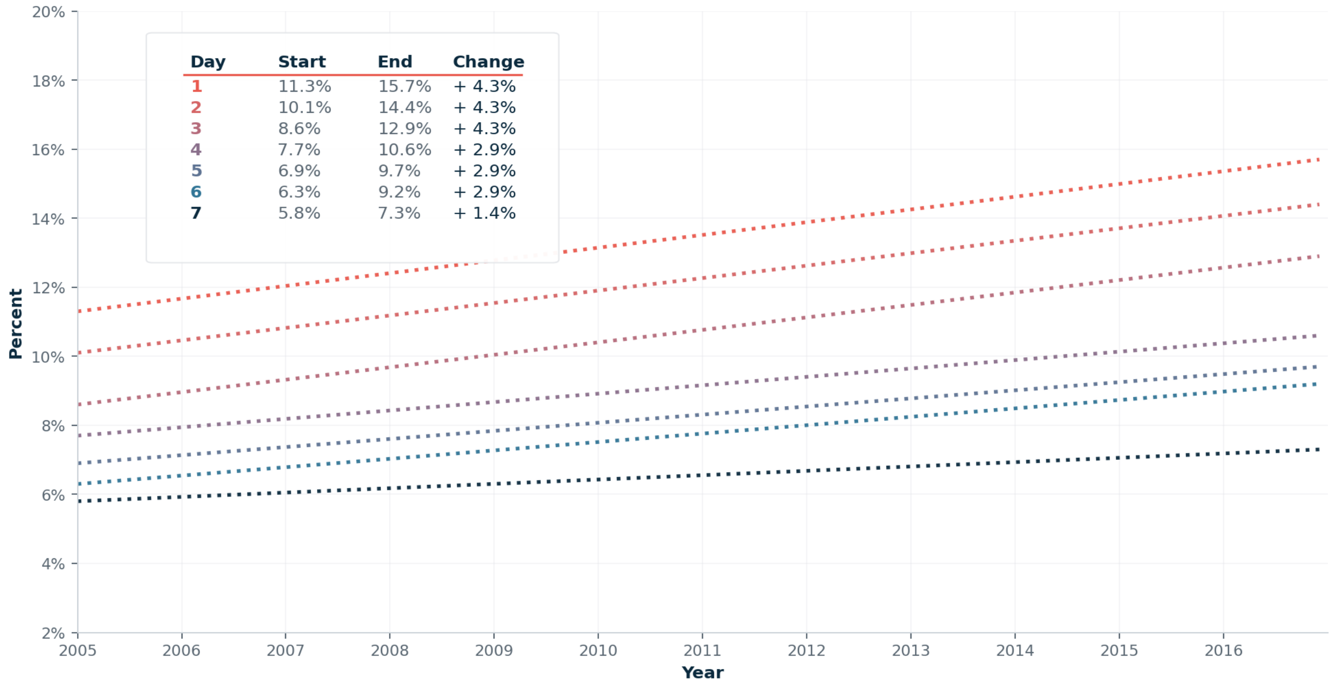

Figure 3 • Day 1–7 Perfect Forecast Percentages by Year, 2005–2016

Results

One-day-out forecasts are accurate, averaging under 3°F error

Figure 4 and Figure 5 show the monthly average error (RMSE) for each provider. For each month, the set of forecast errors are averaged to represent the monthly average RMSE. Each smoothed line represents a different provider. The figures show that accuracy is seasonal but has been steadily improving over the twelve-year period of the study. Today's one-day-out forecasts are extremely accurate, averaging less than ±3°F error.

- **Some forecasts are better than others.** There is considerable spread in the data. The variance between RMSE is notable, suggesting that some providers are clearly performing better than others.
- **Accuracy is cyclical.** Temperature forecasting is easiest in summer when day-to-day variability is the least. Winter forecasts are considerably more challenging because the thermal gradient at the surface of the earth and aloft is much steeper, and the overall flow or progression of weather elements is much faster.
- **Forecast accuracy is improving.** The figure shows a clear trend of decreasing RMSE during the twelve-year period, reflecting improvement in forecasts by all ten providers.

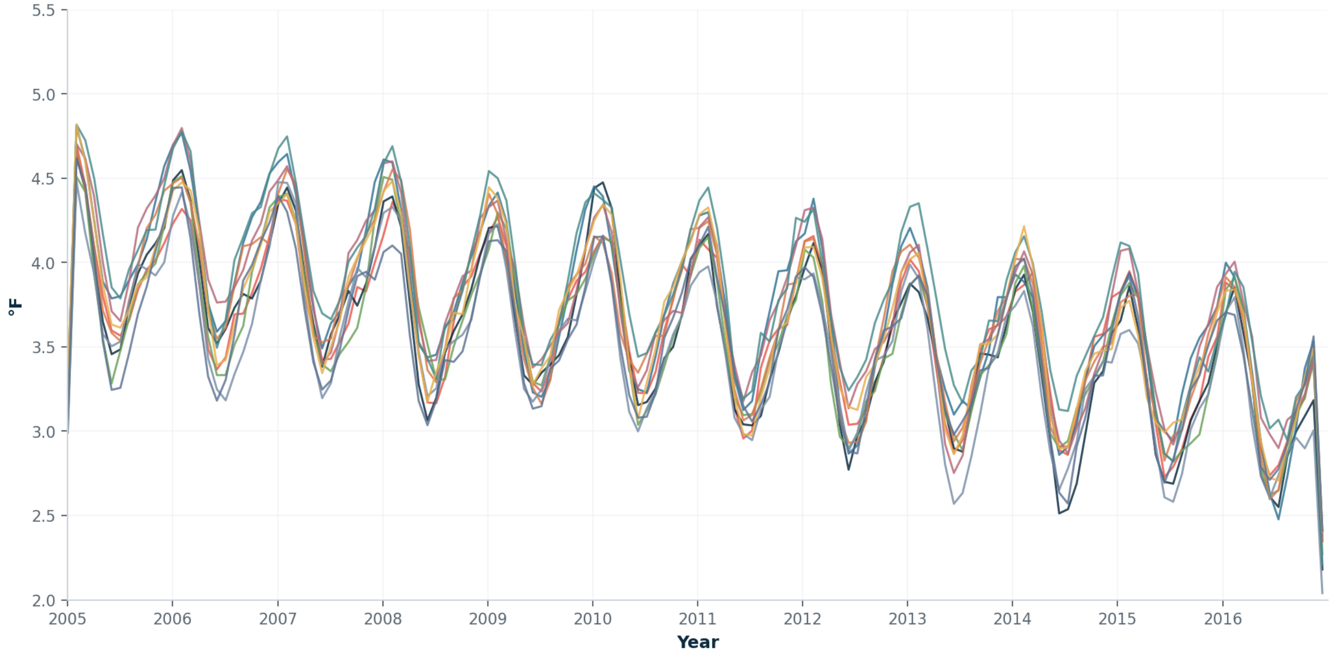

Figure 4 • One-Day-Out Forecast Error by Month by Provider, 2005–2016

Results

A 33% reduction in one-day-out forecast error in twelve years

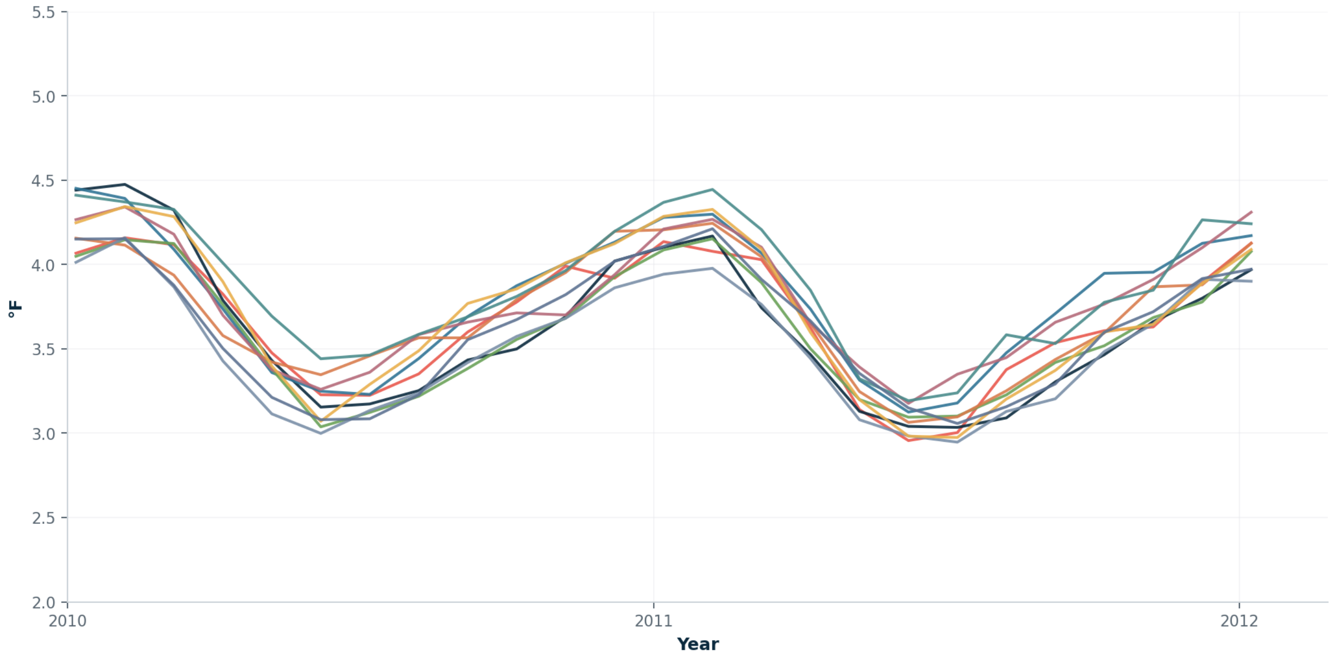

Figure 5 • One-Day-Out Forecast Error by Month by Provider, 2010–2012

Figure 6 shows the minimum, maximum, and average RMSE of one-day-out forecasts for all ten providers by month. The red line depicts the least skillful provider each month and the blue line depicts the most skillful provider each month. The black dotted line is the simple average of the monthly minimum and maximum. The coral dotted line is a computer-generated linear trendline that best expresses the average plotline. Figure 7 shows the skill envelope, average, and trendline for the percentage of one-day-out forecasts within ±3°F of the observation.

- Over the twelve-year study period, forecasts on average became nearly one full degree more accurate. Considering average error is approximately 3°F, this 33% improvement is substantial.
- Consistent with lower RMSE, the fraction of forecasts that were within ±3°F of the observation also improved a similar amount.
- Today, one-day-out forecasts are within ±3°F of the observation more than 80% of the time, whereas in 2005 only 70% were within ±3°F.
- The frequency of forecasts within ±3°F of the observation increased by about one percentage point per year.

Results

One-day-out skill envelope

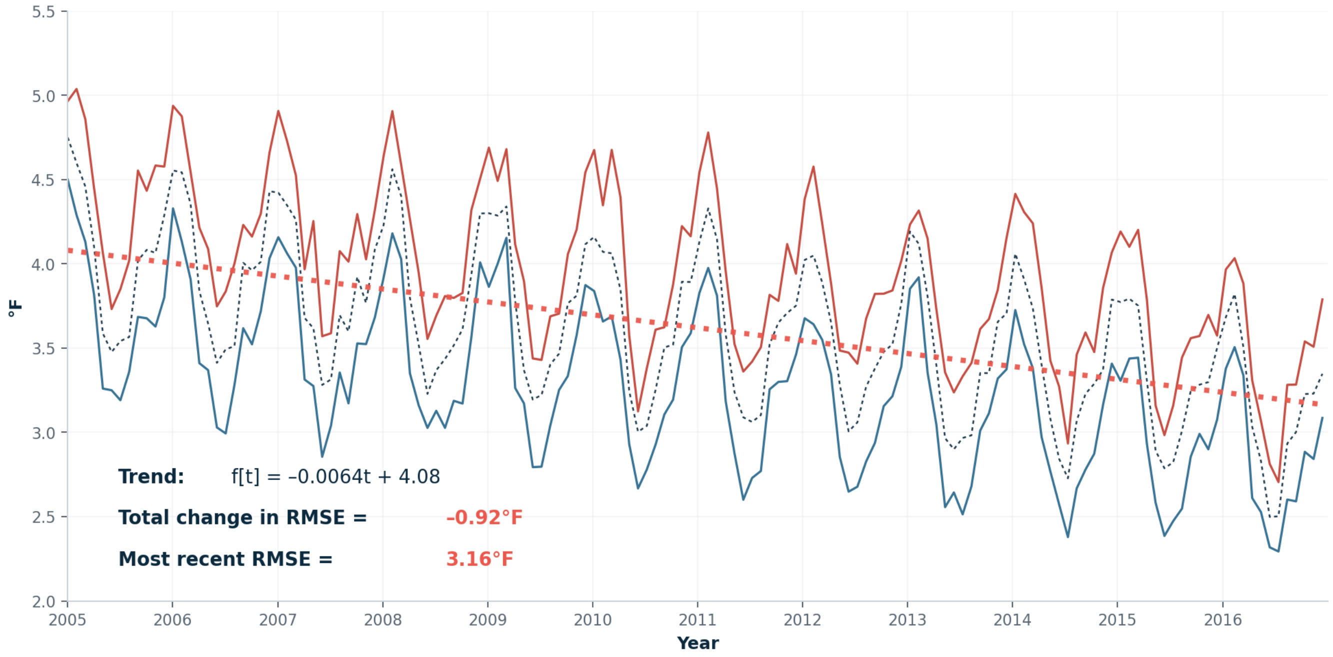

Figure 6 • One-day-out monthly minimum, maximum, and average RMSE from all providers

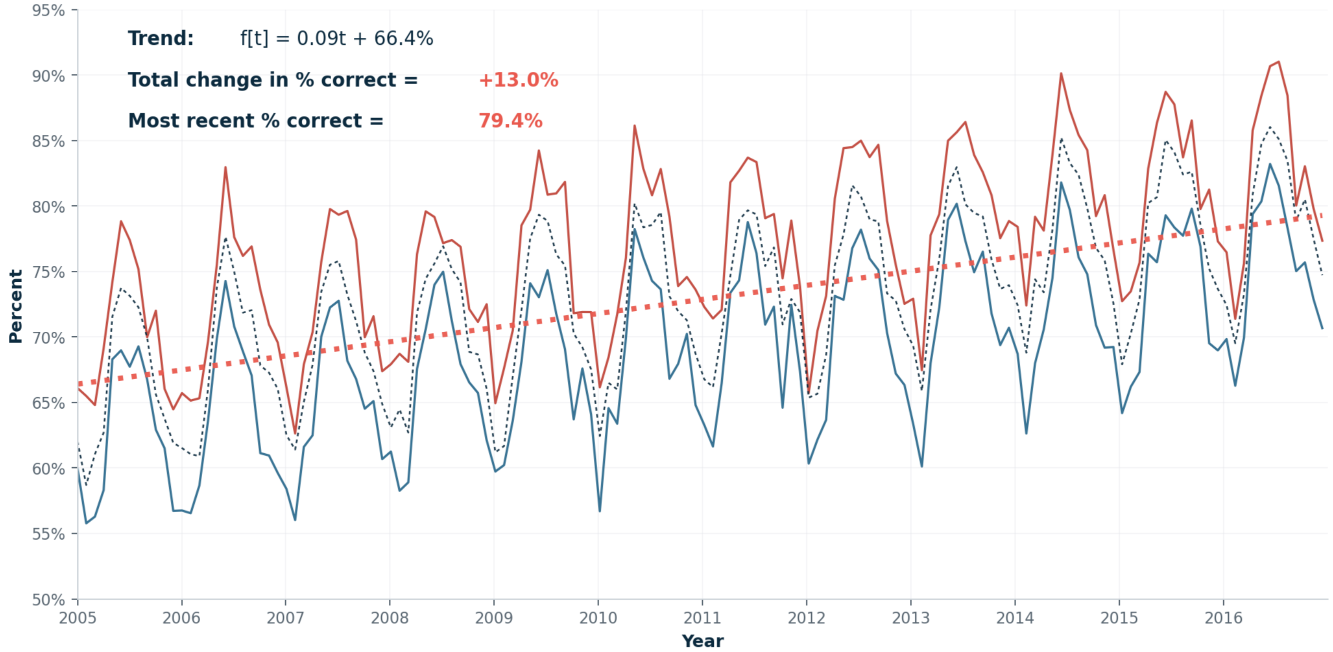

Figure 7 • One-day-out minimum, maximum, and average percent of forecasts within ±3°F from all providers

Results

Five-day-out forecasts have gained the most accuracy

Figure 8 shows the minimum, maximum, and average RMSE of five-day-out forecasts for all ten providers by month. Figure 9 shows the skill envelope, average, and trendline for the percentage of five-day-out forecasts within ±3°F of the observation.

- Five-day-out forecasts improved by 2°F on average, double the improvement of one-day-out forecasts.
- In 2016 the average error was approximately 5°F, improving from 7°F in 2005, an improvement of 40% over the twelve years of the study.
- Five-day-out forecasts within ±3°F increased from 45% to nearly 60% over the study period, becoming nearly as good as one-day-out forecasts were twelve years ago.
- Five-day-out forecast average RMSE was 5°F in 2016. At the beginning of the study, one-day-out forecasts had 4°F error.
- A five-day-out forecast in 2016 is 30% less likely to be within ±3°F than a one-day-out forecast in 2016, but the gap has been declining over the past twelve years.
- The yearly cycle of skill remains prominent. There is greater range or amplitude to the yearly cycle within five-day-out forecasts than one-day-out forecasts.
- If this rate of improvement continues for the next ten years, it is likely that five-day-out high temperature forecasts will approach the same skill as one-day-out forecasts from 2005 — a truly remarkable achievement.

Results

Five-day-out skill envelope

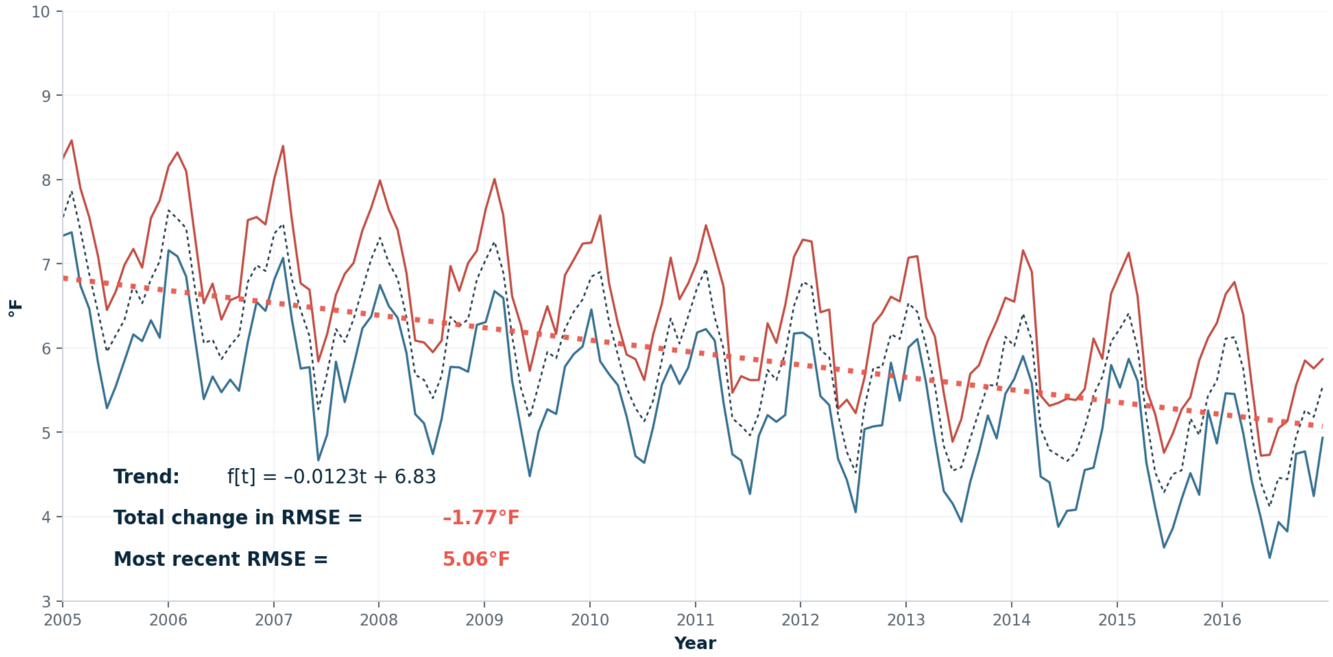

Figure 8 • Five-day-out monthly minimum, maximum, and average RMSE from amongst all providers

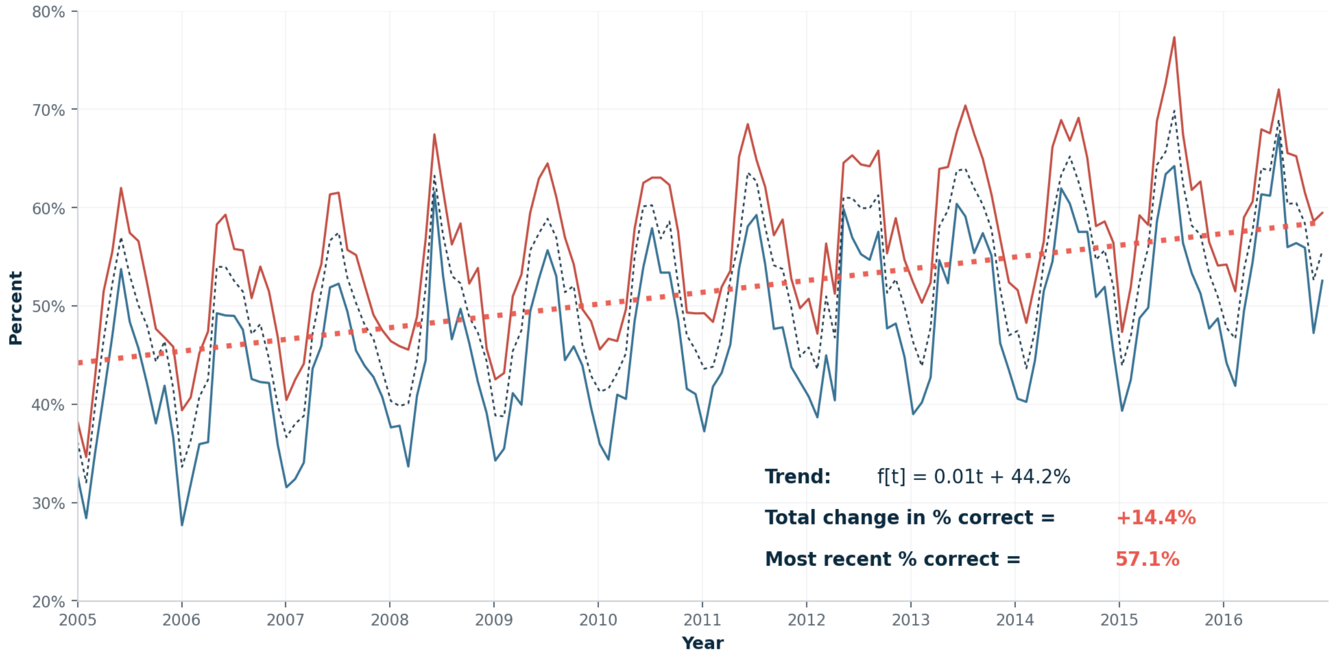

Figure 9 • Five-day-out minimum, maximum, and average percent of forecasts within ±3°F

Results

Nine-day-out forecasts are now just becoming skillful

Figure 10 plots forecast performance (as RMSE) for two providers of nine-day-out forecasts. The green dotted line represents the RMSE of an unskilled climatology forecast, which is a forecast based strictly on 1971–2000 climate normal average high temperature for the forecast date. So, for example, if a nine-day-out forecast is made for Charleston, South Carolina, for April 2, 2013, we can look up the climatology for weather station CHS for April 2, use that as the forecast, and then compare it to the actual observation that occurred that day.

- A climatology forecast has a RMSE approaching 12°F in winter and about 6°F in summer. This qualifies as a bust forecast in winter, and still not very skillful in summer.
- In the early years, we see that the forecaster error lines nearly intersect with climatology, but the difference becomes greater in later years. This reflects improvement in the performance of nine-day-out forecasts against climatology.
- Nine-day-out forecasts of high temperature have some skill over assuming climatologically normal conditions, and therefore have some utility as a prediction.

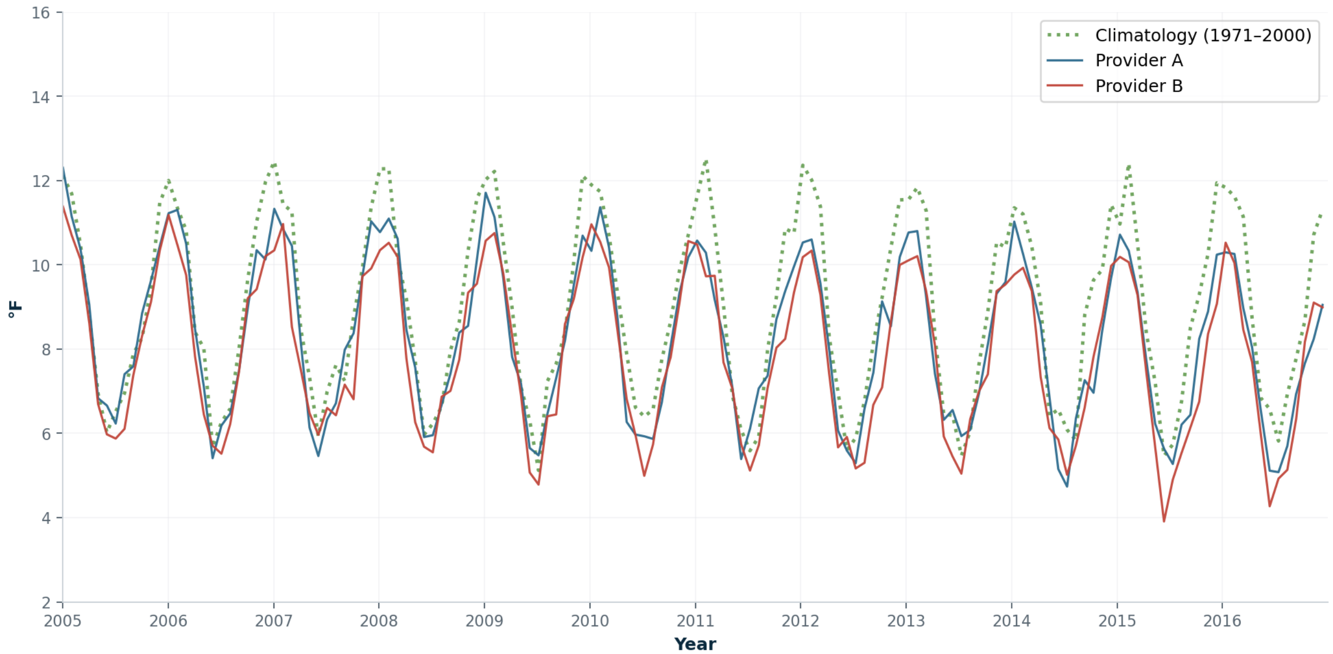

Figure 10 • Day 9 RMSE for two providers and climatology, 2005–2016

Results

Forecasts are generally warmer than actual, but are improving

Bias is measured in terms of mean error, the difference between forecast temperature and actual temperature. This is a measure of systemic warm (positive) or cold (negative) bias in a forecast. Ideally, the average of standard error, or bias, would be near zero, which would indicate that errors are unbiased and tend to even out between forecasts that are too warm and forecasts that are too cool. Figure 11 shows one-day-out average forecast bias by month aggregated for all providers.

- All providers exhibit warm bias for all forecast lengths (one-day-out forecasts are shown).
- Bias appears to be declining. However, since 2011, the per-year improvement is less pronounced. The degree of seasonal variability of bias has improved since 2011.
- Average forecast bias was approximately 0.7°F too warm in 2005, declining to 0.2°F too warm in 2016.

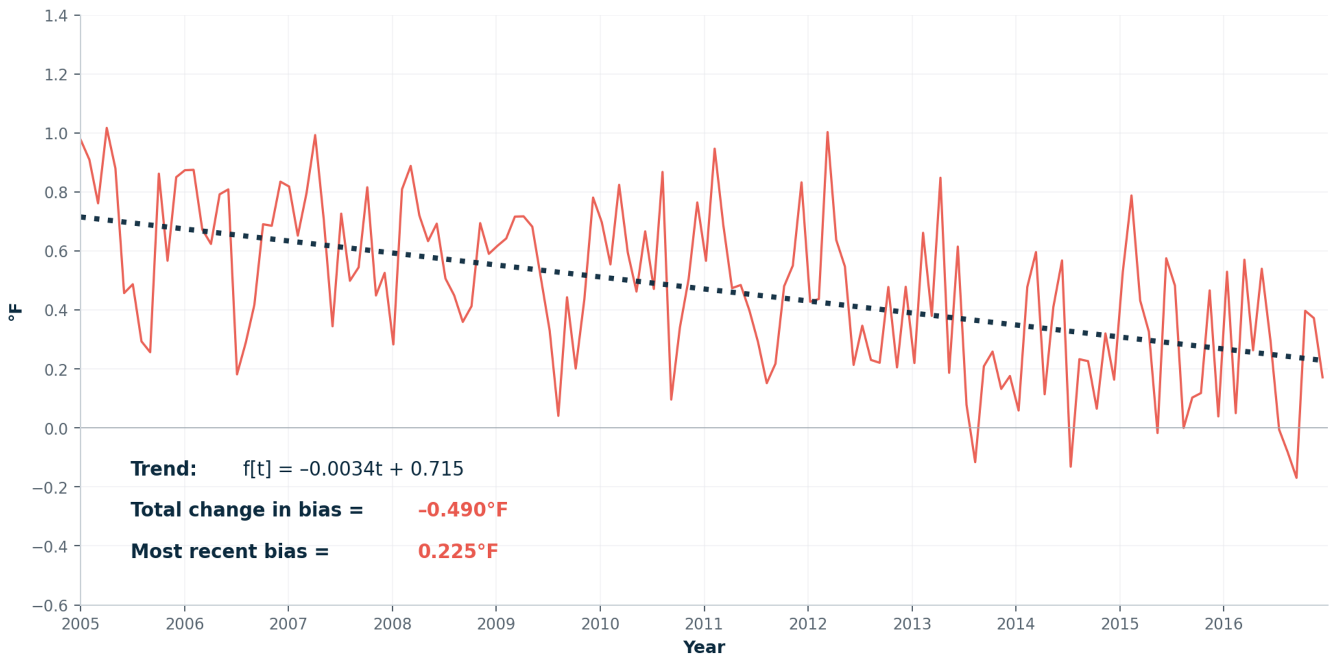

Figure 11 • One-day-out forecast average error of all providers, 2005–2016

About this report

ForecastWatch is the only independent platform where forecast producers benchmark accuracy against anonymized market peers. We verify performance before and after forecast day, isolate station-level issues, and let our customers know exactly where their model stands.

Meteorologists, utilities, and energy companies depend on ForecastWatch's accurate data and analysis. Our data meets the highest standard of scientific inquiry and has been used in several peer-reviewed studies.