ForecastWatch

*Accuracy Defined*

# Long-Term Analysis of Short-Term High Temperature Forecasts

(September 2006 through September 2014)

*By ForecastWatch.com, a Service of Intellovations, LLC*
*December 13, 2014*


**Contact:**

Eric Floehr
*Founder and President*
Intellovations, LLC

508 Grace Drive
Marysville, Ohio 43040
eric@forecastwatch.com
http://www.forecastwatch.com
(614) 440-0130

## Executive Summary

Since September 2006, ForecastWatch.com has been collecting and verifying high temperature forecasts from the public websites of AccuWeather, CustomWeather, Intellicast, the National Weather Service, The Weather Channel, a forecast feed from the NWS National Digital Forecast Database, Weather Underground, and a forecast feed from Schneider Electric. This year, one- and two-day-out high temperature forecasts were evaluated over approximately 800 locations in the United States between September 1, 2013 and August 31, 2014.

A total of almost 4.3 million one- and two-day-out high temperature forecasts were collected from the eight weather forecast providers and compared with daily high temperature observations from the Automated Surface Observation Network. For the eighth straight year, Schneider Electric had the lowest forecast error for the period.

Over the eight years of this study, nearly 35 million forecasts have been collected and compared with observations. Over that eight year period, the weather forecast providers' rankings have remained remarkably consistent. All providers except the National Weather Service website have shown improved accuracy over the eight year period. Schneider Electric has had the greatest improvement in accuracy over eight years with in improvement of nearly one-half of a degree, a 12.5% improvement.

## Why Accurate Temperature Forecasts Are Important

Accurate temperature forecasts are particularly valuable for electric and gas utilities. These weather forecasts are the main driver of their load forecasting, which is used for generation and purchasing decisions. Accurate temperatures pay many benefits including avoiding excessive base loads, reducing spinning reserves and reducing costs overall. Trading and power marketers also greatly benefit from accurate temperature forecasts resulting in smarter market transactions and higher revenues.

Public works departments and state DOTs also benefit from accurate temperature forecasts which are one of the drivers of accurate pavement temperature and frost forecasts. This allows better decision making in crew call-outs, pre-treating roads and overall public safety.

## How Temperature Forecasts Are Evaluated

Temperature forecast accuracy is measured a number of ways. All accuracy calculations begin with taking the forecast high temperature and subtracting the actual observed high temperature. This number is called the error. A forecast that predicts too low a temperature will have a negative error, while a forecast that is too high will have a positive error.

Average absolute error is a measure of the accuracy of temperature forecasts. This measure takes the absolute value of the error of each forecast, so that all errors are positive, and then averages all errors. This is a measure of how far off, on average, the set of forecasts is, regardless if they are too high or too low.

Root-mean-squared (RMS) error takes the square of the absolute error, averages all errors for the set of forecasts, and then takes the square root of the average. With standard absolute error, a forecast that is 2 degrees off is only considered twice as bad as one that is 1 degree off. With RMS error, it is considered four times as bad. Thus, forecasts that are less accurate are more heavily penalized. A forecast that is off by ten degrees is considered one hundred times worse than a forecast that is only one degree off.

# Results of Temperature Forecast Comparison

The one-day-out high temperature forecast RMS errors in degrees Fahrenheit for the period September 1, 2013 through August 31, 2014 are shown in Table 1. For consumers of forecasts who rely on the most accurate forecasts and where temperature differences can significantly change decision-making, RMS error is usually the best accuracy measurement. With RMS, larger forecast errors are penalized much more than small forecast errors. The RMS Error column is the calculated RMS error in degrees Fahrenheit for the period. Rank is the ordered rank of providers for this period (a lower RMS is better).

***One-Day-Out High Temperature Forecast Error for period 9/1/2013-8/31/2014***

| Rank | Provider | RMS Error *(lower is better)* |
|---|---|---|
| 1 | *Schneider Electric* | 3.15 |
| 2 | *The Weather Channel* | 3.52 |
| 3 | *WX Underground* | 3.60 |
| 4 | *Intellicast* | 3.67 |
| 5 | *NDFD* | 3.71 |
| 6 | *CustomWeather* | 3.72 |
| 7 | *AccuWeather* | 3.84 |
| 8 | *NWS Web* | 4.06 |

*Table 1: Results of one year one-day-out high temperature forecast analysis (lower is better)*

The two-day-out high temperature forecast RMS errors are shown in Table 2. The two-day-out results are similar to the one-day-out results, with all providers having the same relative rank as the one-day-out results. Overall, error for two-day-out forecasts are greater than one-day-out forecasts, as expected. On average, there is an approximately 0.38 degree Fahrenheit (0.21 degree Celsius) increase in error between the one- and two-day-out high temperature forecasts.

***Two-Day-Out High Temperature Forecast Error for period 9/1/2013-8/31/2014***

| Rank | Provider | RMS Error *(lower is better)* |
|---|---|---|
| 1 | *Schneider Electric* | 3.59 |
| 2 | *The Weather Channel* | 3.92 |
| 3 | *WX Underground* | 4.00 |
| 4 | *Intellicast* | 4.05 |
| 5 | *NDFD* | 4.08 |
| 6 | *CustomWeather* | 4.11 |
| 7 | *AccuWeather* | 4.22 |
| 8 | *NWS Web* | 4.38 |

*Table 2: Results of one year two-day-out high temperature forecast analysis (lower is better)*

For one-day-out forecasts, the spread between first and last place is 0.91 degrees Fahrenheit (0.51 degrees Celsius), and 0.79 degrees Fahrenheit (0.38 degrees Celsius) for two-day-out forecasts. This is the largest spread in the eight years of the study for one-day-out forecasts, and tied with the largest spread for two-day-out forecasts. For companies where a degree of difference in a forecast can have monetary implications, it is clear that selecting the right forecast provider and continually monitoring their forecasts is important.

## Eight Year Comparison

For the eighth year, Schneider Electric had the lowest RMS error, and this year was one of only two providers who improved their accuracy over last year. While the gap between Schneider Electric and the second-place provider had been decreasing over the previous three years, Schneider Electric reasserted it's lead this year. The difference between Schneider Electric and the second place provider, Weather Underground, was 0.37 degrees Fahrenheit (0.21 degrees Celsius) for one-day-out forecasts, which was the second largest difference in the study period.

Table 3 lists the one-day-out rankings and RMS error for all eight year-long periods. The Weather Underground moved from seventh to third place, moving everyone else down one place. Other than that, rankings remained the same for one-day-out accuracy.

***One-Day-Out High Temperature Forecast Error By Year***

| Year | Schneider Electric | | TWC | | WXU | | Intellicast | | NDFD | | Custom-Weather | | Accu-Weather | | NWS Web | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***2013-2014*** | 3.15 | 1 | 3.52 | 2 | 3.60 | 3 | 3.67 | 4 | 3.71 | 5 | 3.72 | 6 | 3.84 | 7 | 4.06 | 8 |
| ***2012-2013*** | 3.25 | 1 | 3.28 | 2 | 3.75 | 7 | 3.42 | 3 | 3.50 | 4 | 3.67 | 5 | 3.73 | 6 | 3.90 | 8 |
| ***2011-2012*** | 3.34 | 1 | 3.40 | 2 | 3.93 | 5 | 3.48 | 3 | 3.60 | 4 | 4.08 | 8 | 4.06 | 7 | 3.95 | 6 |
| ***2010-2011*** | 3.44 | 1 | 3.70 | 2 | 4.07 | 6 | 3.74 | 4 | 3.72 | 3 | 4.21 | 7 | 4.27 | 8 | 4.02 | 5 |
| ***2009-2010*** | 3.25 | 1 | 3.52 | 2 | 3.75 | 5 | 3.57 | 4 | 3.53 | 3 | 3.84 | 6 | 4.06 | 8 | 3.86 | 7 |
| ***2008-2009*** | 3.45 | 1 | 3.91 | 3 | 4.12 | 5 | 3.97 | 4 | 3.84 | 2 | 4.23 | 7 | 4.31 | 8 | 4.14 | 6 |
| ***2007-2008*** | 3.68 | 1 | 3.83 | 2 | | | 3.89 | 4 | 3.87 | 3 | 4.09 | 6 | 4.29 | 7 | 3.99 | 5 |
| ***2006-2007*** | 3.60 | 1 | 3.71 | 2 | | | 3.77 | 4 | 3.75 | 3 | 4.02 | 6 | 4.18 | 7 | 3.92 | 5 |

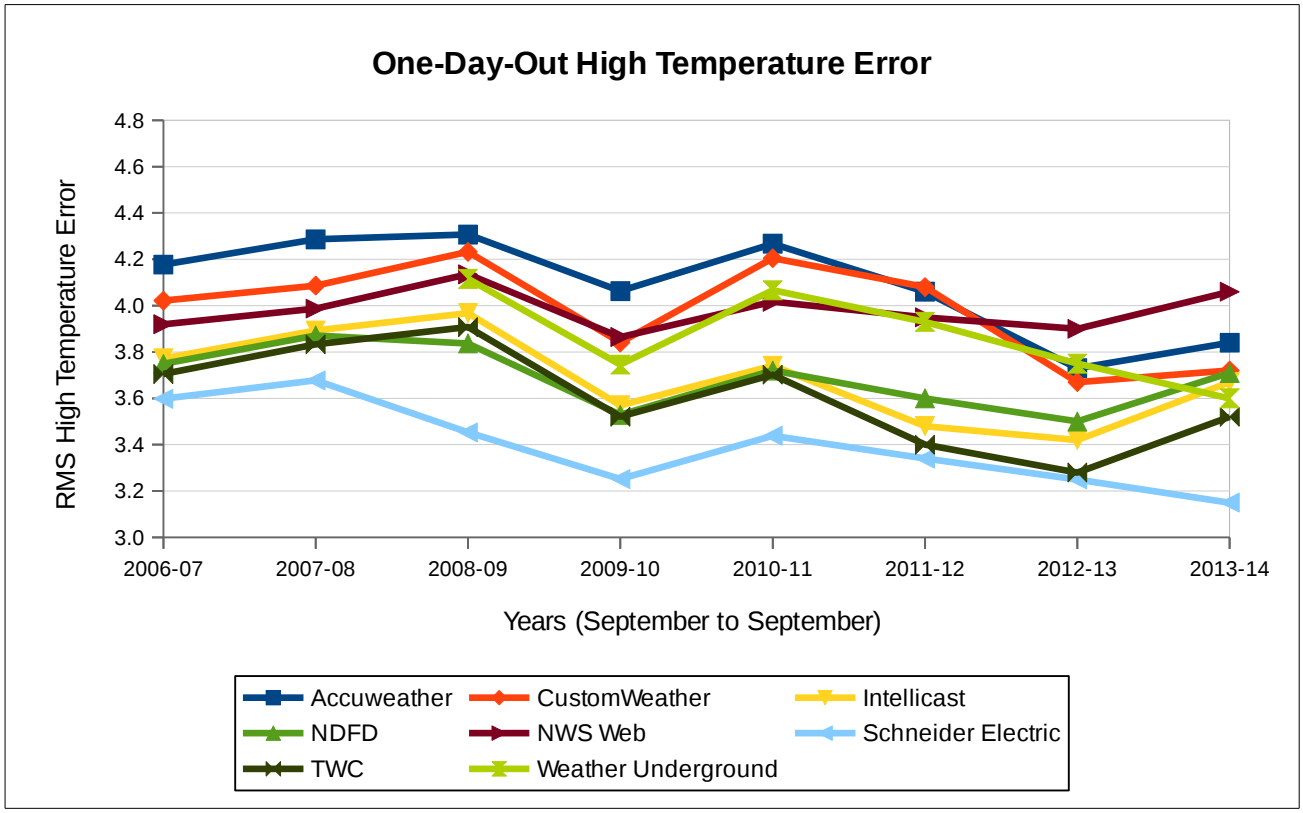
*Table 3: Eight year comparison of one-day-out high temperature forecast error (lower is better)*

Table 4 lists the two-day-out rankings and RMS error for all eight years. Like the one-day-out year-over-year results, Weather Underground moved from seventh to third, moving the other providers down one rank. The difference between Schneider Electric and the second place provider, Weather Underground, was 0.33 degrees Fahrenheit (0.18 degrees Celsius) for two-day-out forecasts, which, like the one-day-out results, was the second largest difference in the study period.

## *Two-Day-Out High Temperature Forecast Error By Year*

| Year | Schneider Electric | | TWC | | WXU | | Intellicast | | NDFD | | Custom-Weather | | Accu-Weather | | NWS Web | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***2013-2014*** | 3.59 | 1 | 3.92 | 2 | 4.00 | 3 | 4.05 | 4 | 4.08 | 5 | 4.11 | 6 | 4.22 | 7 | 4.38 | 8 |
| ***2012-2013*** | 3.70 | 1 | 3.72 | 2 | 4.17 | 7 | 3.83 | 3 | 3.89 | 4 | 4.01 | 5 | 4.13 | 6 | 4.23 | 8 |
| ***2011-2012*** | 3.85 | 1 | 3.88 | 2 | 4.37 | 6 | 3.94 | 3 | 4.03 | 4 | 4.43 | 8 | 4.42 | 7 | 4.35 | 5 |
| ***2010-2011*** | 3.92 | 1 | 3.70 | 3 | 4.45 | 6 | 4.20 | 4 | 4.12 | 2 | 4.57 | 7 | 4.63 | 8 | 4.39 | 5 |
| ***2009-2010*** | 3.70 | 1 | 3.97 | 2 | 4.17 | 5 | 4.00 | 4 | 3.97 | 3 | 4.27 | 7 | 4.45 | 8 | 4.25 | 6 |
| ***2008-2009*** | 3.97 | 1 | 4.34 | 3 | 4.59 | 6 | 4.43 | 4 | 4.33 | 2 | 4.67 | 7 | 4.76 | 8 | 4.57 | 5 |
| ***2007-2008*** | 4.20 | 1 | 4.32 | 2 | | | 4.39 | 4 | 4.37 | 3 | 4.50 | 6 | 4.67 | 7 | 4.47 | 5 |
| ***2006-2007*** | 4.13 | 1 | 4.20 | 2 | | | 4.29 | 3 | 4.29 | 4 | 4.54 | 6 | 4.64 | 7 | 4.44 | 5 |

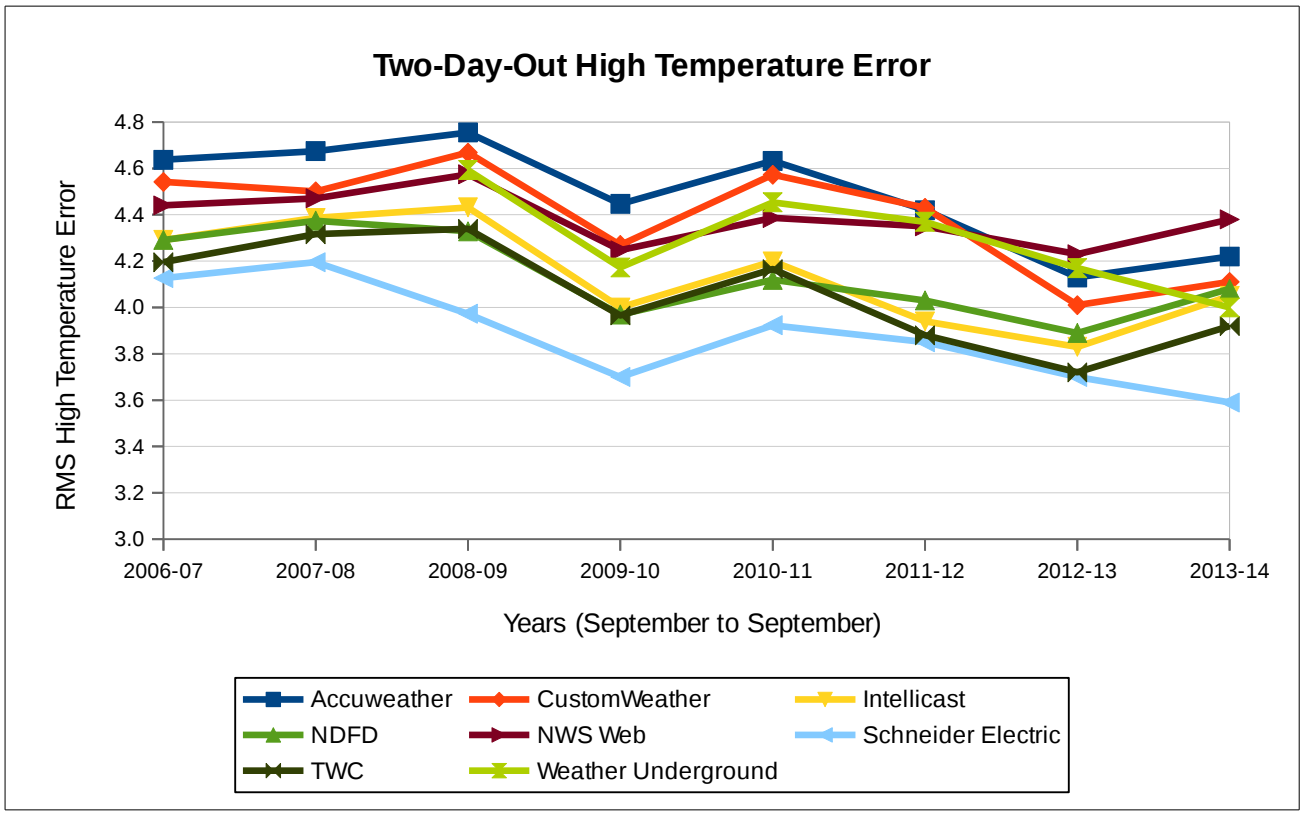
*Table 4: Eight year comparison of two-day-out high temperature forecast error (lower is better)*




*Graph 1: Eight year graph of one-day-out high temperature forecast error (lower is better)*

The two line graphs illustrate the changes in high temperature forecast error over the eight year study period and the accuracy differences between providers. Graph 1 shows the eight year history of high temperature forecast RMS error for one-day-out forecasts, while Graph 2 shows the eight year history for two-day-out forecasts. The y-axis of both graphs do not begin at zero, in order to highlight differences and trends in accuracy over the seven-year period, however the vertical scale of both graphs is identical.

It is apparent from the graphs that Schneider Electric has improved its accuracy for both one- and two-day-out forecasts for four years in a row, as has Weather Underground. The other providers all showed decreased accuracy in lock-step from the previous year. A trend of overall increasing accuracy of high temperature forecasts is however still apparent in the eight year trends.




*Graph 2: Eight year graph of two-day-out high temperature forecast error (lower is better)*

## Methodology of the Comparison

Daily high temperature forecasts were collected from each provider starting at 22:00 UTC (6pm Eastern Standard Time) and continuing until all forecasts are collected. For each location, forecasts from all providers were collected at the exact same time. Forecasts from AccuWeather were collected from http://www.AccuWeather.com premium site, which provides the same daily forecasts as their free site, but is advertising-free and provides additional capabilities. Intellicast was collected from http://www.intellicast.com, The Weather Channel (TWC) forecasts were from http://www.weather.com, the National Weather Service forecasts from http://www.weather.gov, and CustomWeather's forecasts from http://www.myforecast.com. The NDFD forecasts were collected using a SOAP interface at http://www.weather.gov/ndfd/. Weather Underground forecasts were collected from http://www.wunderground.com.

A one-day-out high temperature forecast is the forecast for the next day, whereas the two-day-out forecast is for the day after that. For example, for a forecast collected on January 1, 2014, the one-day-out high temperature forecast would be the forecast for January 2, 2014, and the two-day-out forecast would be the forecast for January 3, 2014.

| Provider | Number of Forecasts | Percent of Possible Forecasts |
|---|---|---|
| ***AccuWeather*** | 270,882 | 92.0% |
| ***CustomWeather*** | 269,835 | 91.6% |
| ***Intellicast*** | 271.301 | 92.1% |
| ***NDFD*** | 264,714 | 89.9% |
| ***NWS Web*** | 261,227 | 88.7% |
| ***Schneider Electric*** | 271,301 | 92.1% |
| ***The Weather Channel*** | 271,272 | 92.1% |
| ***WX Underground*** | 269,807 | 91.6% |

*Table 5: Number of one-day-out forecasts analyzed and percent of possible, by provider*

| Provider | Number of Forecasts | Percent of Possible Forecasts |
|---|---|---|
| ***AccuWeather*** | 270,103 | 91.7% |
| ***CustomWeather*** | 269,051 | 91.3% |
| ***Intellicast*** | 270,518 | 91.8% |
| ***NDFD*** | 263,941 | 89.6% |
| ***NWS Web*** | 260,471 | 88.4% |
| ***Schneider Electric*** | 270,518 | 91.8% |
| ***The Weather Channel*** | 270,490 | 91.8% |
| ***WX Underground*** | 269,014 | 91.3% |

*Table 6: Number of two-day-out forecasts analyzed and percent of possible, by provider*

For this study, the 24-hour high temperatures as reported by the observation stations in the ASOS/AWOS observation network maintained by the National Weather Service and the Federal Aviation Administration were used. These observations were collected from the Quality-Controlled Local Climatic Data (QCLCD) product from the National Climatic Data Center (NCDC).

The Root-Mean-Squared (RMS) error was calculated by subtracting the high temperature observation from the high temperature forecast, and that value was squared. The square root of the average of all the squared errors for each year was calculated, and that is the value shown in the report. Each yearly period includes forecasts collected September 1 through August 31 the following year.

There were around 270,000 forecasts used for each provider and year. The percent of possible forecasts collected and compared is less than 100% because of invalid forecasts, problems in collecting forecasts successfully, and days in which observations were not available for a particular site.

Table 5 shows the number of forecasts collected and compared for each provider for one-day-out forecasts, and Table 6 shows the same for two-day-out forecasts. Differences in the forecast count are due to a number of factors. First, invalid forecasts are removed. Second, occasionally a provider's website or feed would be off-line or not be complete due to network issues, production issues in the provider's forecast creation, or other issues. For example, the relatively lower number of NWS web forecasts is primarily due to quality issues (invalid forecasts) and website availability issues (the weather.gov site was unavailable or could not provide a forecast). Finally, ASOS/AWOS stations are down and don't provide data for a day or two every few months due to maintenance. Overall, around 91% of the possible forecasts and observations that could be compared for each provider were able to be compared.

## About ForecastWatch.com

ForecastWatch is the nation’s premier weather forecast monitoring and assessment company. A full-service, technology consulting firm, ForecastWatch compiles weather forecasts and observations at more than 1,200 locations around the world, including the U.S., Canada, Europe, South America, Central America, Africa and Asia Pacific. ForecastWatch also maintains a historical database of over 400 million weather forecasts from a number of providers.

ForecastWatch data and analysis has been used by meteorologists, utilities and energy companies, the agriculture industry, futures traders, and others whose business success depends on being right about the weather. Our data meets the highest standard of scientific inquiry, and has been used in several peer-reviewed studies, including a paper published in the Monthly Weather Review. In 2003, ForecastWatch.com released the largest public weather forecast accuracy study undertaken to that point.

ForecastWatch services have been used to evaluate weather forecast providers, improve decision-making where weather forecasts are used as input, improve weather forecasts by providing useful feedback, compare weather forecast performance between providers, educate customers with unbiased reporting, and improve the quality of weather forecast websites.