ForecastWatch

Accuracy Defined

Six-Month Analysis of High Temperature Forecasts

(July 2013 through December 2013)

By ForecastWatch.com, a Service of Intellovations, LLC
January 23, 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

ForecastWatch collected high temperature forecasts from eight commercial forecast providers and the U.S. National Weather Service over a period of six-months from July 1, 2013 through December 31, 2013. More than 11 million forecasts were collected for over 900 locations in the United States and Europe. These forecasts were compared with 7 a.m. to 7 p.m. (MOS) high temperature observations in those locations.

Global Weather Corporation had the lowest high temperature forecast error overall for both U.S. and Europe. Global Weather Corporation also had the least error for all days-out ranges for Europe, and for two of the three days-out groupings in the U.S. MeteoGroup had the lowest high temperature forecast error in the 4-6 day-out range for the U.S.

The rankings were relatively stable across both regions. Global Weather Corporation, MeteoGroup, and The Weather Channel were the top three providers in both the U.S. and Europe for the 1-3 and 4-6 day-out forecasts, with Foreca replacing MeteoGroup in the 7-9 day-out range and overall. World Weather Online and AccuWeather were at the bottom of the accuracy rankings for both regions.

High Temperature Scoring Methodology

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.

Rank is determined by sorting on RMS error, with the provider with the least error and thus the most accurate forecasts ranked first. This ranking is statistically significant with 99% confidence. In the few cases where there is not 99% confidence, i.e. confidence intervals overlap, it is noted and rank is considered a tie.

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, and so on. For example, for a forecast collected on October 1, 2013, the one-day-out high temperature forecast would be the forecast for October 2, 2013, and the two-day-out forecast would be the forecast for October 3, 2013, and the nine-day-out forecast would be the forecast for October 10, 2013.

Analysis of Forecast Providers' High Temperature Accuracy

This study looked at both the U.S. and Europe for overall 1-9 day out accuracy. It also broke the forecasts into three groups by the forecast number of days out over the six month period from July 1, 2013 through December 31, 2013. Providers that did not provide forecasts for the full forecast range in a group were excluded from that group. Only locations in which all forecasting companies or organizations provided forecasts were used. A total of 775 stations in the U.S. and 166 in Europe were forecast by all providers in the study.

Results For The United States

Table 1 shows the accuracy results for the U.S. The table is ordered by overall accuracy ranking, then by 1-3 day-out forecasts for those excluded from the overall ranking. For the most part, the overall rankings held for the other groupings. Global Weather Corporation had the lowest overall error and the lowest error for 1-3 day-out and 7-9 day-out forecasts, while MeteoGroup had the lowest for 4-6 day-out forecasts.

A total of 1.808 degrees Fahrenheit separated the best from the worst provider overall, while 1.752, 2.089, and 1.786 degrees Fahrenheit separated the best from the worst for 1-3, 4-6, and 7-9 day-out forecasts respectively. The amount of difference between best and worst is significant, and important in industries where accuracy matters, such as the energy and transportation industries, where even a half-degree improvement in accuracy results in noticeable improvements to the bottom line. The rankings are also statistically significant to at least a 99% confidence level.

Three commercial providers provided higher accuracy forecasts than the National Weather Service. While all providers utilize government-sponsored models like GFS, NAM, and the ECMWF, all incorporate them in different ways, have additional modeling and analysis layers, and use human expertise differently. It is not true that all forecast providers simply repackage and reuse the NWS forecast.

Relative rankings across days out are fairly consistent, with a few exceptions. CustomWeather did relatively poorly in the 4-6 day-out and 7-9 day-out rankings compared with their 1-3 day-out performance. Foreca did relatively better in the 4-6 and 7-9 day-out range compared to their 1-3 day-out error.

Provider	1-3 Day-Out		4-6 Day-Out		7-9 Day-Out		1-9 Day-Out	
	RMS Err	Rank	RMS Err	Rank	RMS Err	Rank	RMS Err	Rank
Global Weather Corporation	3.347	1	5.071	2	7.046	1	5.351	1
MeteoGroup	3.484	2	5.026	1				
The Weather Channel	3.520	3	5.240	3	7.258	2	5.533	2
National Weather Service	3.736	4	5.494	4				
Foreca	3.959	7	5.670	5	7.349	3	5.809	3
Weather Underground	3.912	6	5.821	6	7.915	4	6.082	4
AccuWeather	4.094	8	6.035	7	7.964	5	6.213	5
CustomWeather	3.834	5	6.088	8	8.300	6	6.317	6
World Weather Online	5.099	9	7.115	9	8.832	7	7.159	7

Table 1: High Temperature RMS Error in Degrees Fahrenheit for U.S. Locations

Results for Europe

Table 2 shows the accuracy results for Europe. As with the U.S. table, it is ordered by overall accuracy ranking. Similar to U.S. forecasts, rankings between the days-out groupings were quite consistent. The overall rankings between the U.S. and Europe were also quite similar. A total of 1.753, 1.780, 2.042, and 1.600 degrees Fahrenheit separated the most accurate provider from the least accurate for the overall, 1-3, 4-6, and 7-9 day-out groupings. This is similar to the U.S. numbers. Overall, temperature forecasting in Europe was more accurate, with providers having had lower error for the same groups compared with the U.S.

Global Weather Corporation had the lowest overall error, and thus the highest accuracy, for all day-out groupings. MeteoGroup was in second in the 1-3 and 4-6 day-out groups, while Foreca took second for 7-9 day-out. The Weather Channel was a close third in all ranges. In Europe, CustomWeather did relatively better than in the U.S., while World Weather Online had the least accurate forecasts in both the U.S. and Europe. Compared with the 1-3 day-out group, AccuWeather and Weather Underground switched places, and in the 7-9 day-out range, Foreca jumped over The Weather Channel to take second place. Note that with the exception

of CustomWeather and Weather Underground in the 1-3 day-out group, the ranking is statistically significant to the 99% confidence level.

Provider	1-3 Day-Out		4-6 Day-Out		7-9 Day-Out		1-9 Day-Out	
	RMS Err	Rank	RMS Err	Rank	RMS Err	Rank	RMS Err	Rank
Global Weather Corporation	2.705	1	3.869	1	5.328	1	4.093	1
MeteoGroup	2.912	2	3.923	2				
The Weather Channel	3.019	3	4.155	3	5.489	3	4.325	2
Foreca	3.200	4	4.280	4	5.397	2	4.372	3
CustomWeather	3.313	5*	4.528	5	6.160	4	4.792	4
Weather Underground	3.324	5*	4.754	7				
AccuWeather	3.637	7	4.610	6	6.821	5	5.176	5
World Weather Online	4.485	8	5.911	8	6.928	6	5.846	6

Table 2: High Temperature RMS Error in Degrees Fahrenheit for European Locations
** - difference not statistically significant at the 99% confidence level*

Forecast Collection Methodology

Daily high temperature forecasts for locations in the U.S. were collected from each provider starting at 22:00 UTC (6 p.m. Eastern Standard Time) and continued until all forecasts were collected. For European locations, forecasts were collected starting at 16:00 UTC (6 p.m. Eastern European Standard Time). For each location, forecasts from all providers were collected at the exact same time, with the exception of Global Weather Corporation forecasts, which were collected in a bulk feed at the start of the forecast collection time.

Forecasts were considered valid if they were complete (i.e. they contained a high and low temperature forecast), and if they passed both manual and automated audits. These audits checked for out-of-bounds values and other indicators that suggested the forecast should be marked as invalid. Forecasts that were simply bad were not considered invalid. However, forecast issues caused by system bugs or delivery problems (such as a -32768 degree high temperature) were declared invalid.

Collecting AccuWeather Forecasts

Forecasts from AccuWeather were collected from http://www.accuweather.com using the Premium, ad-free 15-day forecast page. For the U.S., the location parameters used to retrieve the specific forecasts were of the form {number}_PC. For Europe, a “city, country” name was used for the location parameter. If that resulted in ambiguity or multiple forecast locations, an AccuWeather-specific location ID, which is a number corresponding to the correct location, was used. Temperature forecasts were collected in degrees Fahrenheit.

Collecting CustomWeather Forecasts

Forecasts from CustomWeather were collected from http://www.myforecast.com using the expanded 15 day forecast page. The location parameter used to retrieve the specific forecasts for both the U.S. and Europe was

the observation station ICAO code, or SYNOP code if the location did not have an ICAO code. Temperature forecasts where collected in degrees Fahrenheit.

Collecting Foreca Forecasts

Forecasts from Foreca were collected from http://www.foreca.com using the ten-day forecast page. For the U.S., the location parameter used was the city and state of the observation location. For Europe, the parameter consisted of the country and city of the observation location. Forecasts were collected in degrees Celsius and converted to Fahrenheit.

Collecting Global Weather Corporation Forecasts

Forecasts from Global Weather Corporation were collected from a CSV file retrieved at the beginning of each regions' forecast collection. Locations were identified by ICAO code, and the MaxTemperature7to7 parameter presented on hour 0 of the given day was used as the daily high temperature forecast for that location. Forecasts were collected in degrees Fahrenheit.

Collecting MeteoGroup Forecasts

Forecasts from MeteoGroup were collected from http://www.weathercast.co.uk using the 0-6 day-out chart presented on the main forecast page for each location. While MeteoGroup provides a 7-14 day-out forecast, it was not collected due to a current single-forecast-page limitation of ForecastWatch collection. The location parameters were MeteoGroup-specific location codes collected by using the city search function. Where the city search returned multiple results or the search parameters were ambiguous, the MeteoGroup-specific location code was determined manually. Forecasts were collected in degrees Celsius and converted to Fahrenheit.

Collecting National Weather Service Forecasts (U.S. Only)

Forecasts from the U.S. National Weather Service were collected from the National Digital Forecast Database using the SOAP interface documented here http://graphical.weather.gov/xml/. Forecasts were requested using the latitude and longitude of the observation station location. Forecasts were collected in degrees Fahrenheit.

Collecting The Weather Channel Forecasts

Forecasts from The Weather Channel were collected from http://www.weather.com using the ten-day forecast page. For both the U.S. and Europe, the location parameter used to retrieve specific forecasts was the ICAO code or SYNOP of the observation station. Temperature forecasts were collected in degrees Fahrenheit.

Collecting Weather Underground Forecasts

Forecasts from Weather Underground were collected from http://www.wunderground.com using the ten-day forecast provided under Weather Underground's “BestForecast” forecast. For both the U.S. and Europe, the location parameter used to retrieve specific forecasts was the ICAO code or SYNOP of the observation station. Temperature forecasts were collected in degrees Fahrenheit.

Collecting World Weather Online Forecasts

Forecasts from World Weather Online were collected from http://www.worldweatheronline.com using their premium weather API. Latitude and longitude of the observation station were parameters to the API query to retrieve the specific forecasts for both U.S. and European locations. Temperature forecasts were collected in degrees Fahrenheit.

Observation Collection and Calculation Methodology

Observational data for U.S. locations was procured from the primary ASOS weather observation network. The data were quality controlled by National Climatic Data Center (NCDC) systems and personnel prior to delivery to ForecastWatch via the Quality-Controlled Local Climatic Data (QCLCD) product data set. This product consisted of both hourly and daily observation parameters.

Observational data for European locations were sourced from a number of primary sources. These observations were collected and quality controlled by National Climatic Data Center (NCDC) systems and personnel prior to deliver to ForecastWatch via the Integrated Surface Database (ISD) product. This product consisted of hourly observations.

The maximum temperature from the 7 a.m. to 7 p.m. local time hourly observations was used to construct the 7 a.m. to 7 p.m. (MOS) high temperature. No attempt to curve fit or otherwise determine an intra-hour temperature estimate was performed.

Provider	1-3 Day-Out		4-6 Day-Out		7-9 Day-Out		1-9 Day-Out	
	Count	Pct	Count	Pct	Count	Pct	Count	Pct
AccuWeather	385,083	91.0%	378,581	91.0%	372,170	91.0%	1,135,834	91.0%
CustomWeather	385,099	91.0%	378,601	91.0%	372,179	91.0%	1,135,879	91.0%
Foreca	380,805	90.0%	374,659	90.0%	368,602	90.1%	1,124,066	90.0%
Global Weather Corporation	385,185	91.0%	378,689	91.0%	372,266	91.0%	1,136,140	91.0%
MeteoGroup	385,110	91.0%	378,620	91.0%				
National Weather Service	357,606	84.5%	330,152	79.3%				
The Weather Channel	385,120	91.0%	378,622	91.0%	372,202	91.0%	1,135,944	91.0%
Weather Underground	384,801	90.9%	378,322	90.9%	371,243	90.7%	1,134,366	90.9%
World Weather Online	381,288	90.1%	375,008	90.1%	368,759	90.1%	1,125,055	90.1%

Table 3: Valid Forecasts Used in Accuracy Calculations for U.S. Locations

Analysis of Forecast and Observation Collection

Table 3 shows the counts of forecasts used in the accuracy calculations for U.S. locations. Table 4 shows the counts of forecasts for European locations. The percentage represents the fraction of a theoretical maximum number of forecasts that could have been incorporated, assuming all forecasts and all observations for every date in the study range were valid.

The number of forecasts used was generally around 90% of the theoretical maximum. The largest component contributing to lower actual forecasts compared with the theoretical maximum was missing observations. Missing observations were usually due to station maintenance, station malfunction, or issues with delivery. The difference between the number of forecasts between providers was due exclusively to missing or invalid forecasts from a provider. For example the 91% of AccuWeather, CustomWeather, Global Weather Corporation, MeteoGroup, and The Weather Channel forecasts compared to the 84.5% of the National Weather

Service forecasts was due exclusively to missing or invalid forecasts from the NWS. The NWS SOAP service is much less reliable, and down more often, than the websites of the major commercial providers.

Finally, the reason that the theoretical maximum of the 1-3 day-out forecasts was higher than the 4-6 day-out, and the 4-6 day-out count higher than the 7-9 day-out, was due to January observations not being used for the study. This report evaluated forecasts created from July 1, 2013, through December 31, 2013. Forecasts created on December 31, 2013 had forecasts for dates in January 2014. However, at the time of this report, the January observational data was not complete (there is generally a delivery lag due to the quality control efforts used on the observational data) so any forecasts for dates in January 2014 were not used.

In the final analysis, a total of 9,378,772 forecasts for the U.S. were used for this study, which is 90.0% of the theoretical maximum number of forecasts that could have been used. For European locations, a total of 1,787,903 forecasts were used, for an average of 90.7% of the theoretical maximum.

Provider	1-3 Day-Out		4-6 Day-Out		7-9 Day-Out		1-9 Day-Out	
	Count	Pct	Count	Pct	Count	Pct	Count	Pct
AccuWeather	81,512	89.9%	80,119	89.9%	78,536	89.9%	240,167	89.9%
CustomWeather	83,052	91.6%	81,576	91.5%	79,920	91.2%	244,548	91.4%
Foreca	82,439	91.0%	81,034	90.9%	79,456	90.7%	242,929	90.8%
Global Weather Corporation	82,654	91.2%	81,179	91.1%	79,520	90.7%	243,353	91.0%
MeteoGroup	83,014	91.6%	81,565	91.5%				
The Weather Channel	83,125	91.7%	81,648	91.6%	79,985	91.3%	244,758	91.5%
Weather Underground	83,014	91.6%	81,552	91.5%				
World Weather Online	82,481	91.0%	81,044	90.9%	79,478	90.7%	243,003	90.9%

Table 4: Valid Forecasts Used in Accuracy Calculations for European Locations

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.