ForecastWatch

Accuracy Defined

Analysis of One- to Nine-Day-Out High Temperature Forecasts for U.S., Europe, and Asia-Pacific

(Calendar Year 2014)

By ForecastWatch.com, a Service of Intellovations, LLC
March 18, 2015

Contact:

Eric Floehr
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 the full calendar year 2014, or January 1, 2014 through December 31, 2014. Nearly 24 million forecasts were collected for over 950 locations in the United States, Europe, and the Asia-Pacific regions. 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 U.S., Europe, and Asia-Pacific. Global Weather Corporation also had the least error for all days-out ranges for U.S., Europe, and Asia-Pacific, though was in a statistical tie for first with MeteoGroup for the 4-6 day-out range for the U.S. and Asia-Pacific.

The rankings were similar between regions, with Global Weather Corporation and MeteoGroup at the top of the rankings. Weather Underground and The Weather Channel came third and fourth in the U.S., but Foreca leaped them internationally for third place in Europe and Asia-Pacific. CustomWeather and World Weather Online were at the bottom of the accuracy rankings in all three 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 each absolute error, averages all errors for the set of forecasts, and then takes the square root of the average. Samples are weighted linearly in estimating absolute error but as squares in estimating RMS error. Thus, a set of forecasts that show greater variance are penalized more than forecasts with consistent error. For example, three forecasts with an error of one, four, and ten degrees would have an average absolute error of 5 degrees. Those same three forecasts would have a RMSE of 6.24 degrees. However a set of three forecasts all with five degrees error will have both a MAE and RMSE of 5 degrees.

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. If we were able to collect an infinite amount of forecasts, we would be able to determine error perfectly. However we are sampling each provider's total set of forecasts for a region. Thus our calculation of error is not the actual error, but a very good estimate of that error. This estimate gets better the more forecasts we use. Since this is an estimate we also calculate a range where the actual error is likely located, and the range that we use is one that 99 times out of 100 the actual error is within that range. 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, 2014, the one-day-out high temperature forecast would be the forecast for October 2, 2014, and the two-day-out forecast would be the forecast for October 3, 2014, and the nine-day-out forecast would be the forecast for October 10, 2014.

Analysis of Forecast Providers' High Temperature Accuracy

This study looked the U.S., Europe, and the Asia-Pacific region for overall 1-9 day out accuracy. It also broke the forecasts into three groups by the forecast number of days out over the full year from January 1, 2014 through December 31, 2014. 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 766 locations in the U.S., 130 locations in Europe, and 58 locations in Asia-Pacific 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, with the exception of CustomWeather dropping from second-to-last to last place in the 7-9 day-out forecast range. Global Weather Corporation had the lowest overall error and the lowest error for 1-3 day-out, 4-6 day-out, and 7-9 day-out forecasts. In the 4-6 day-out range, Global Weather Corporation was statistically tied with MeteoGroup for first place at the 99% confidence level.

A total of 1.233 degrees Fahrenheit separated the best from the worst provider overall, while 1.156, 1.363, and 1.508 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 with exceptions noted by asterisks.

Four 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.

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	3.484	1	5.329	1*	7.549	1	5.703	1
MeteoGroup	3.625	2	5.342	1*				
Weather Underground	3.648	3*	5.477	3	7.630	2	5.818	2
The Weather Channel	3.644	3*	5.499	4	7.697	3	5.854	3
National Weather Service	3.848	5	5.799	5*				
Foreca	3.925	6	5.818	5*	7.888	4	6.097	4
AccuWeather	3.976	7	6.118	7	8.678	5	6.548	5
CustomWeather	4.017	8	6.431	8	9.057	7	6.822	6
World Weather Online	4.640	9	6.692	9	8.828	6	6.936	7

Table 1: High Temperature RMS Error in Degrees Fahrenheit for U.S. Locations
** - difference not statistically significant at the 99% confidence level*

Results for Europe

Table 2 shows the accuracy results for Europe. As with the U.S. table, it is ordered by overall accuracy ranking. The rankings between overall and days-out groupings were completely consistent. The overall rankings between the U.S. and Europe were also quite similar, with the exception that Foreca did notably better, jumping from fourth place overall in the U.S. to second in Europe. A total of 1.763, 1.730, 1.844, and 1.816

degrees Fahrenheit separated the most accurate provider from the least accurate for the overall, 1-3, 4-6, and 7-9 day-out groupings. These ranges are greater than the U.S. accuracy range. However, 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 as well as overall. 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, Weather Underground, and AccuWeather were all statistically tied for fourth place in the 1-3 day-out range, while Foreca, Weather Underground, and The Weather Channel were all statistically tied for third in the 4-6 day-out range. Like the U.S., CustomWeather and World Weather Online performed poorly.

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	2.783	1	3.961	1	5.347	1	4.164	1
MeteoGroup	2.936	2	4.028	2				
Foreca	3.074	3	4.303	3*	5.549	2	4.426	2
Weather Underground	3.175	4*	4.299	3*				
The Weather Channel	3.176	4*	4.305	3*	5.658	3	4.496	3
AccuWeather	3.202	4*	4.375	6	6.280	4	4.790	4
CustomWeather	3.279	7	4.725	7	6.493	5	5.009	5
World Weather Online	4.513	8	5.805	8	7.163	6	5.927	6

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

Results for Asia-Pacific

Table 3 shows the accuracy results for Europe. Unlike the U.S. and Europe, there is somewhat less consistency between the days-out groupings. However, the top two and bottom two are the same for all three regions. Global Weather Corporation had the lowest RMS error overall, and also for each days-out grouping. MeteoGroup was in second place in the 1-3 day-out grouping, and statistically tied with Global Weather Corporation in the 4-6 day-out range. A total of 2.685, 3.151, 3.002, and 2.119 degrees Fahrenheit separated the most accurate provider from the least accurate for the overall, 1-3, 4-6, and 7-9 day-out groupings. There is much greater difference between providers in Asia-Pacific than the other regions. Overall, accuracy tended to be better than the U.S., and the same or better than Europe.

Like the U.S. and Europe, Global Weather Corporation had the lowest overall error, and thus the highest accuracy, for all day-out groupings. Like the U.S., MeteoGroup was in second in the 1-3 and 4-6 day-out groups, but statistically tied with Global Weather Corporation for first place in the 3-6 day-out group. Foreca was third in the 1-3 and 4-6 day-out groups, while taking second in the 7-9 and overall days-out groups. As with the other two regions, CustomWeather and World Weather Online lagged.

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	2.886	1	3.701	1*	4.698	1	3.843	1
MeteoGroup	2.983	2	3.752	1*				
Foreca	3.261	3	4.085	3	4.978	2	4.167	2
The Weather Channel	3.624	5*	4.335	5	5.206	3	4.436	3
AccuWeather	3.600	5*	4.499	7	5.665	4	4.665	4
Weather Underground	3.584	5*	4.251	4				
CustomWeather	3.477	4	4.455	6	6.176	5	4.834	5
World Weather Online	6.037	8	6.703	8	6.817	6	6.528	6

Table 3: High Temperature RMS Error in Degrees Fahrenheit for Asia-Pacific 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 locations in the Asia-Pacific region, forecasts were collected starting at 08:00 UTC (6 p.m. Australia Eastern 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 and Asia-Pacific 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	789,853	94.2%	789,806	94.2%	791,722	94.4%	2,371,381	94.2%
CustomWeather	785,615	93.7%	785,658	93.7%	787,603	93.9%	2,358,876	93.7%
Foreca	782,745	93.3%	782,509	93.3%	784,118	93.5%	2,349,372	93.4%
Global Weather Corporation	788,023	93.9%	787,971	93.9%	789,910	94.2%	2,365,904	94.0%
MeteoGroup	790,014	94.2%	789,971	94.2%				
National Weather Service	780,849	93.1%	727,208	86.7%				
The Weather Channel	785,635	93.7%	785,698	93.7%	787,535	93.9%	2,358,868	93.7%
Weather Underground	785,504	93.6%	785,368	93.6%	785,873	93.7%	2,356,745	93.7%
World Weather Online	777,717	92.7%	777,485	92.7%	779,351	92.9%	2,334,553	92.8%

Table 4: Valid Forecasts Used in Accuracy Calculations for U.S. Locations

Analysis of Forecast and Observation Collection

Table 4 shows the counts of forecasts used in the accuracy calculations for U.S. locations. Table 5 shows the counts of forecasts for European locations. Finally, Table 6 shows the counts of forecasts used for locations in the Asia-Pacific region. 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 above 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.

In the final analysis, a total of 19,583,741 forecasts for the U.S. were used for this study, which is 93.4% of the theoretical maximum number of forecasts that could have been used. For European locations, a total of 2,899,020 forecasts were used, for an average of 92.6% of the theoretical maximum. Finally, for locations in the Asia-Pacific region, a total of 1,284,446 forecasts were used, for an average of 91.9% of the theoretical

maximum. All told, a total of 23,767,207 forecasts were used in this study out of a theoretical maximum of 25,498,170 forecasts, which is 93.2% of all possible forecasts.

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	132,666	93.2%	132,678	93.2%	132,682	93.2%	398,026	93.2%
CustomWeather	131,906	92.7%	132,020	92.7%	132,159	92.8%	396,085	92.7%
Foreca	131,247	92.2%	131,248	92.2%	131,218	92.2%	393,713	92.2%
Global Weather Corporation	133,048	93.5%	133,049	93.5%	133,049	93.5%	399,146	93.5%
MeteoGroup	131,780	92.6%	131,760	92.6%				
The Weather Channel	130,443	93.2%	132,661	93.2%	132,661	93.2%	397,982	93.2%
Weather Underground	130,443	91.6%	130,475	91.7%				
World Weather Online	129,864	91.2%	129,885	91.2%	129,861	91.2%	389,610	91.2%

Table 5: Valid Forecasts Used in Accuracy Calculations for European Locations

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	58,871	92.7%	58,869	92.7%	58,869	92.7%	176,609	92.7%
CustomWeather	58,355	91.9%	58,333	91.8%	58,430	92.0%	175,118	91.9%
Foreca	58,164	91.6%	58,201	91.6%	58,128	91.5%	174,493	91.6%
Global Weather Corporation	58,759	92.5%	58,759	92.5%	58,760	92.5%	176,278	92.5%
MeteoGroup	57,910	91.2%	57,903	91.2%				
The Weather Channel	58,759	92.5%	58,761	92.5%	58,762	92.5%	176,282	92.5%
Weather Underground	58,610	92.3%	58,581	92.2%				
World Weather Online	57,571	90.6%	57,552	90.6%	57,539	90.6%	172,662	90.6%

Table 6: Valid Forecasts Used in Accuracy Calculations for Asia-Pacific 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 500 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.