

Extended Forecast Verification at the Weather Forecast Office at Nashville, Tennessee

Mark A. Rose, Scott Dickson, and Darrell Massie

1. Introduction

Just how good are the extended forecasts we issue? This is a question common to many meteorologists. With current verification programs focused on just the first few periods of the forecast, little (if any) research has been conducted on the "extended portion" (days 3 through 5) of zone forecast products issued by National Weather Service Offices across the country.

In this study, extended forecasts issued by the Weather Forecast Office (WFO) at Nashville (BNA) are evaluated for the period 1 November 1999 through 31 October 2000, a one-year period. Verification methods specific to this study are described in section 2. Only precipitation and temperature components were analyzed. Hopefully, more offices will undertake similar studies with the hope of finding out "just how good we are," and with the hope of identifying seasonal trends that may help in refining extended forecasts in the future. In addition, methods that may help improve forecasts will be discussed.

2. Methodology

Daily extended forecasts contained in the afternoon zone forecast product were collected for Davidson County, Tennessee. (Nashville is located in Davidson County.) When compiling verification, the authors used the values recorded in a "daily verification log," which is broken into 12-hour periods (night and day) for periods 5 and 6. For days 4 and 5, the preliminary local climatological data forms (F6's) were used.

Precipitation and temperature components were loaded into a Quattro Pro spreadsheet. Each extended forecast was broken into 4 groups: period 5, period 6, day 4, day 5 (table 1). For each period/day's precipitation forecast, a "0" was entered when no precipitation was mentioned. A "1" was entered if precipitation was mentioned. The forecast temperature range was also included. For instance, if the period 5 forecast was for lows in the lower to mid 60s, then the forecast temperature range was entered as "60-67." If the period 6 forecast was for highs in the 80s, then the forecast temperature range was entered as "80-89." Period 5 therefore contains a precipitation and minimum temperature forecast. Period 6 contains a precipitation and maximum temperature forecast. Days 4 and 5 contain precipitation, minimum temperature, and maximum temperature forecasts.

Forecast Period	Begin Time	End Time
Period 5	1800 CST (Day 2)	0600 CST (Day 3)
Period 6	0600 CST (Day 3)	2400 CST (Day 3)

Day 4	0000 CST	2400 CST
Day 5	0000 CST	2400 CST

When entering the observed data into the spreadsheet, as with the forecast, a "0" was entered for a particular period or day if no measurable precipitation occurred, and a "1" was entered if measurable precipitation occurred. The observed temperature(s) were also added for each period.

a. Precipitation

Three traditional verification statistics are used to analyze the accuracy of precipitation forecasts: the false alarm ratio (FAR), probability of detection (POD), and critical success index (CSI). (See table 2.) It was decided to use these three statistics for one reason. The "null" case happens most of the time, and tends to skew raw verification numbers. By using FAR, POD, and CSI, only verification from those cases when precipitation was forecast and/or measured is considered. This is discussed in greater detail in section 3.

Table 2. Precipitation Forecast Verification Matrix		
	Precipitation was measured	Precipitation was not measured
Precipitation was forecast	Hit	False alarm
Precipitation was not forecast	Miss	Null

The FAR considers all forecasts of precipitation. If precipitation was forecast, but subsequently did not occur, that forecast is considered a "false alarm." The FAR is therefore the ratio of false alarms to the total number of precipitation forecasts.

For the POD, only periods when precipitation was measured are used. The POD is another measure of the forecast accuracy of precipitation events that accounts for hits and misses. It is a ratio of the number of "wet" forecasts for those periods when measurable precipitation fell and the total number of periods when precipitation was measured.

The CSI simply combines these two into one statistic.

Equations used to calculate POD, FAR, and CSI, which are presented in section 3, are listed below.

FAR = false alarms/total number of precipitation forecasts

POD = hits when precipitation was measured/days when precipitation was measured

CSI = hits/(hits + misses + false alarms)

b. Temperatures

If the observed temperature for a period or day fell within the forecast range, the temperature forecast was considered to have verified. For instance if the observed low was 62 when the forecast was for lows in the lower to mid 60s, then the forecast verified. Conversely, if the high temperature was 91 when the forecast was for highs in the 80s, the forecast was not verified, and would be assigned a forecast error of -2 degrees, since high end of the forecast range fell 2 degrees below the observed value (underforecast).

3. Results

Again, verification scores for precipitation and temperatures were calculated a number of ways. Precipitation forecast verification will be addressed first.

a. Precipitation

Table 3 shows precipitation forecast verification results categorized by forecast period and month, and includes FAR, POD, and CSI tabulations.

Table 3. False Alarm Ratio (FAR), Probability of Detection (POD), and Critical Success Index (CSI)																									
Mo nth	Period 5						Period 6						Day 4						Day 5						
	FA R	PO D	C SI	Hi ts	Mis ses	Fals e Alar ms	FA R	PO D	C SI	Hi ts	Mis ses	Fals e Alar ms	FA R	PO D	C SI	Hi ts	Mis ses	Fals e Alar ms	FA R	PO D	C SI	Hi ts	Mis ses	Fals e Alar ms	
Nov	0%	67%	67%	2	1	0	33%	60%	43%	3	2	2	43%	67%	44%	4	2	3	50%	29%	22%	2	5	2	
Dec	62%	100%	38%	5	0	8	45%	86%	50%	6	1	5	27%	100%	73%	8	0	3	44%	71%	45%	5	2	4	
Jan	30%	88%	64%	7	1	3	75%	43%	19%	3	4	9	64%	40%	24%	4	6	7	82%	20%	11%	2	8	9	
Feb	33%	100%	67%	6	0	3	69%	57%	25%	4	3	9	38%	73%	50%	8	3	5	55%	45%	29%	5	6	6	
Mar	36%	78%	54%	7	2	4	58%	63%	33%	5	3	7	59%	64%	33%	7	4	10	64%	45%	25%	5	6	9	
Apr	47%	89%	50%	8	1	7	67%	50%	25%	5	5	10	29%	67%	53%	10	5	4	63%	19%	14%	3	13	5	
May	71%	57%	24%	4	3	10	59%	78%	37%	7	2	10	65%	55%	27%	6	5	11	67%	45%	24%	5	6	10	
Jun	47%	100%	53%	8	0	7	44%	100%	56%	9	0	7	38%	100%	63%	10	0	6	50%	78%	44%	7	2	7	
Jul	87%	50%	12%	2	2	13	63%	88%	35%	7	1	12	52%	91%	45%	10	1	11	56%	67%	36%	8	4	10	
Aug	80%	100%	20%	3	0	12	57%	75%	38%	6	2	8	57%	67%	35%	6	3	8	89%	13%	6%	1	7	8	
Sep	56%	100%	44%	4	0	5	20%	80%	67%	8	2	2	33%	44%	36%	4	5	2	33%	44%	36%	4	5	2	

	%	%	%				%	%	%				%	%	%				%	%	%						
Oct	75%	50%	20%	1	1	3	100%	0%	0%	0	1	5	86%	25%	10%	1	3	6	67%	50%	25%	2	2	4			
Year	57%	84%	40%	57	11	75	57%	72%	36%	63	26	86	49%	68%	41%	78	37	76	61%	43%	26%	49	66	76			

Note that overall POD values were quite high for the first three forecast periods (84%, 72%, and 68%), but dropped to 43% on day 5. Conversely, the FAR exceeded 50% in 3 of the 4 forecast periods. Little, if any, seasonal trends can be drawn from the data.

As aforementioned, null cases were not considered, since they occur most of the time. In fact, during the year of study, there were 249 days with no measureable precipitation, or 68% (table 4). In other words, if forecasters had never included precipitation in any of the extended forecasts during the year of study, they would have been correct 68% of the time. Even though a 68% verification rate for the extended periods would have been excellent at face value, such forecasts would have been of little value. This explains the reason only hits, misses, and false alarms were considered in the precipitation forecast verification.

Parameter	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Year
Days	23	23	18	19	22	15	17	21	22	22	19	28	249
Percentage	77	74	58	66	71	50	45	70	71	71	63	90	68

b. Temperatures

Verification of the temperature forecasts proved even more complicated, since forecast ranges differ frequently. Forecasts with different forecast ranges cannot be compared directly. For instance, a forecast of "lows in the 60s" would have a much better chance of verifying than a forecast of "lows in the mid 60s" since the first example encompasses a range of 10 degrees (60-69), while the second example encompasses a range of 5 degrees (63-67). Therefore, the forecasts were subdivided according to forecast range in order to facilitate a more equitable comparison. Table 5 gives the results of the temperature forecast verification comparison. Note that the large majority of cases in each period fall into either the 5 or 6 degree forecast range.

The average forecast error was also computed for each category. Here, the absolute value of all individual forecast errors were summed and divided by the number of cases. These values do not indicate any trends of over- or underforecasting. For instance, a forecast error of 2.4 indicates that, on average, the observed temperature was 2.4 degrees outside the forecast range. As expected, the verification rates increase and average errors decrease toward larger forecast ranges and earlier periods, whereas verification rates decrease and average errors increase toward smaller forecast ranges and later periods.

Indeed, when using a discreet forecast range of 6 degrees or less, the verification was often less than 40%. This raises an important question. Do these results suggest we should use larger

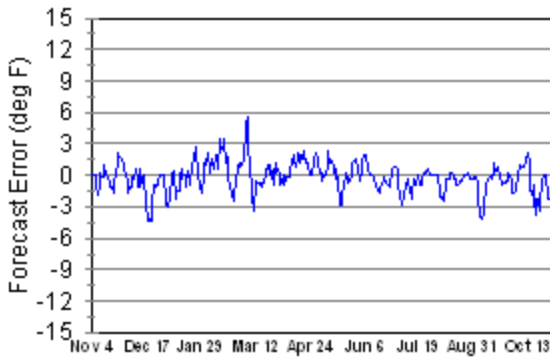
temperature ranges whenever possible? Perhaps this is a question forecasters should consider when trying to determine what degree of specificity to use for extended periods. Based on this study, the tendency to become specific with temperature ranges in the extended forecast is probably something to be avoided, simply because to do so implies a forecast accuracy that is not realistic in the extended periods.

Table 5. Temperature Verification

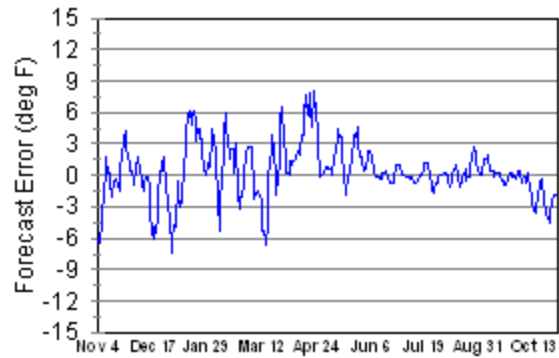
Range	Period 5 Lows		Period 6 Highs				Day 4				Day 5							
	Lows		Highs				Lows		Highs		Lows		Highs					
	Cases	Verification	Avg Error	Cases	Verification	Avg Error	Cases	Verification	Avg Error	Cases	Verification	Avg Error	Cases	Verification	Avg Error			
4	17	35%	2.4	20	5%	5.8	13	23%	5.4	18	0%	6.5	17	24%	4.4	20	10%	5.6
5	190	44%	2.2	169	39%	2.9	159	40%	2.6	147	35%	3.4	146	34%	3.6	153	34%	3.3
6	110	43%	1.8	90	50%	2.0	107	45%	2.6	98	49%	2.1	103	39%	2.6	85	40%	3.1
7	12	75%	0.8	16	69%	2.0	12	50%	1.6	15	40%	3.8	12	42%	2.3	12	33%	1.8
8	2	0%	1.5	12	58%	1.2	3	33%	1.0	18	56%	1.3	1	0%	3.0	14	36%	2.3
9	1	100%	0.0	3	100%	0.0	1	100%	0.0	4	50%	1.0	3	0%	4.0	3	0%	7.7
10	29	48%	1.7	51	49%	2.5	64	50%	2.2	59	37%	2.5	77	49%	2.5	72	39%	3.7
11	0	-	-	1	0%	2.0	1	0%	5.0	2	50%	1.0	1	100%	0.0	2	50%	1.0
12	1	100%	0.0	1	100%	0.0	2	100%	0.0	2	50%	1.5	2	50%	2.0	2	100%	0.0
13	1	100%	0.0	0	-	-	1	0%	1.0	0	-	-	1	100%	0.0	0	-	-

The forecast errors were also analyzed seasonally in an effort to isolate any periods of prolonged over- or underforecasting. Here, daily forecast errors were first plotted for each forecast period. A 5-day moving average was then used in order to reduce the effects of day-to-day variations. In other words, each plot shown in figures 1-6 represents the average forecast error over 5 days, ending that day. Although all forecasts were used in these graphs, regardless of forecast range, the representations do show some interesting trends.

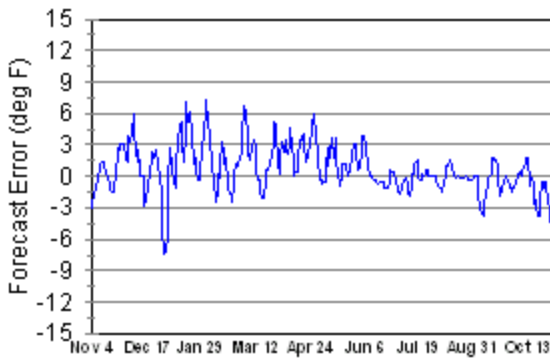
Period 5 Low Temperature Verification
5-Day Moving Average



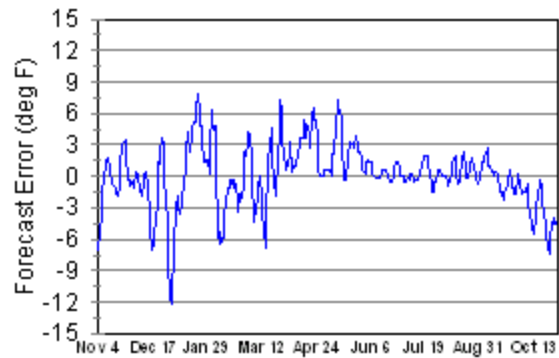
Period 6 High Temperature Verification
5-Day Moving Average



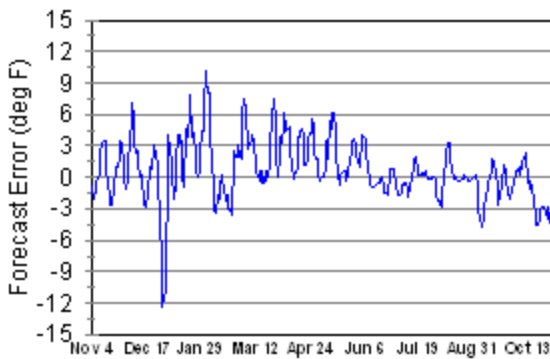
Day 4 Low Temperature Verification
5-Day Moving Average



Day 4 High Temperature Verification
5-Day Moving Average



Day 5 Low Temperature Verification
5-Day Moving Average



Day 5 High Temperature Verification
5-Day Moving Average



Figures 1-6. Temperature forecast errors throughout the year of study.

A period of significant underforecasting is observed toward the end of December, which was subsequently followed by a period of overforecasting throughout much of January. A prolonged period of overforecasting was also observed during the spring, whereas the month of October, near the end of the study period, represents a period of underforecasting. (Note how forecast errors become more amplified toward the later periods.)

Three points that must be addressed here involve model bias, limited data, and climatological variance. Traditionally, forecasters become more reliant on model guidance during later forecast periods. Although no specific conclusions can be drawn from this study, as model guidance values were not tracked, if real-time model biases are found during a particular period, this recognition can be used to adjust future forecasts and diminish the effects of such biases.

In addition, this study was limited to data for only one year. As such, the effects of climatological variance must be seriously considered. It is assumed that a month which exhibits a particularly high absolute departure from normal (i.e., extended periods of above or below normal temperatures) will probably be associated with poorer forecast accuracy simply due to the natural tendency of forecasters to predict conditions that are not extreme. With a data set of only one year, periods exhibiting a high absolute departure will not be "averaged out." As an example, the mean temperature each day during the period October 14-31 (a period represented by persistent underforecasting) was on average 9.6°F above normal.

4. Effects of Changes to the Extended

A potentially contentious issue common among operational meteorologists deals with "changing the forecast." A tangential component of this study will address just that -- do we improve or worsen the extended forecast with subsequent changes? Here, extended forecasts issued by each midnight shift were checked for changes to the forecasts issued the previous afternoon. Only changes to the precipitation components will be studied here (i.e., a "dry" forecast changed to a mention of precipitation, or precipitation completely removed from a "wet" forecast). The results are shown in table 6.

Table 6. The Effects of Changing the Extended Forecast for Precipitation				
Period	Cases	Verification of Original Forecast	Verification of Amended Forecast	Net Change
Period 5	17	41%	59%	+18%
Period 6	18	33%	72%	+39%
Day 4	14	43%	50%	+7%
Day 5	13	46%	54%	+8%

Two conclusions may be drawn from these results: 1) because of the limited number of cases (changes), forecasters seem to exercise prudence in amending the extended forecasts, and 2) the changes that are made do tend to improve the forecasts.

5. Conclusions

Results from a year-long extended forecast verification study at Nashville are presented. Precipitation forecasts were verified using false alarm ratio, probability of detection, and critical success index. POD's were fairly high for the first 3 periods, ranging between 68% and 84%, with FAR's just above 50%. Scores were lower for the last period.

The verification of temperature forecasts becomes a bit complicated, since temperature forecasts go beyond the yes/no methodology used for precipitation forecasts, and cover ranges from 4 to 13 degrees. Not only were the temperature results subdivided by forecast period, but also by forecast range. As expected, verification rates increase toward earlier periods and larger forecast ranges. Verification scores decrease toward later forecast periods and smaller ranges. Also, temperature forecast errors were graphed for the entire year of study (employing a 5-day moving average). A period of underforecasting was noted toward the end of December, followed by a period of overforecasting during January. A prolonged period of overforecasting also occurred during spring, and the month of October represents a period of underforecasting.

Changes to the afternoon extended precipitation forecasts by the following midnight forecaster were quite limited, but did show significant improvements when done.

A logical extension to this study would be to see how these results compare with actual model predictions. Did forecasters actually show improvement over model forecasts for days 3-5? This would also be of interest to forecasters, giving them an idea of how model biases can be isolated and compensated for. Regardless, verification of the extended forecast is a tough issue, because of the various temperature ranges used, and because forecast errors tend to be quite large (when compared to near term forecast verification). Also, precipitation forecasts must be simplified to a yes or no, as probability of precipitation is not used, and qualifying terms, such as "likely" and "scattered" are rarely mentioned.

Acknowledgements

The authors thank Steven Vasiloff, Radar Meteorologist, NSSL/NWS WR-SSD, for his final review and numerous suggestions, Dan Smith, NWS SR-SSD, for his thorough review and helpful suggestions, Henry Steigerwaldt, Science-and-Operations Officer, WFO BNA, for his assistance during this project, and for his review of the manuscript, and Mike Girodo, Lead Forecaster, WFO BNA, for his review.

Last Updated: May 21, 2001

Author: [Mark Rose](#)