9.15 USES OF REAL-TIME VERIFICATION TO IMPROVE TERMINAL AERODROME FORECASTS

David R. Manning
Steven A. Amburn
James M. Frederick
NOAA/National Weather Service Forecast Office
Tulsa, OK

1. INTRODUCTION

Improving Terminal Aerodrome Forecasts (TAFs) has been identified as a goal of *VISION 2005, National Weather Service Strategic Plan for Weather, Water, and Climate Services 2000-2005* (U.S. Department of Commerce, 1999). And, although forecasters have always worked to issue the best possible TAFs, they have received few tools to assist in their effort. Verification is one tool that both forecasters and managers can use to help improve the final TAF product and also to identify forecasters for training which may be having difficulty.

Real-time verification can be available and presented in a way to help forecasters effectively educate themselves and learn how to become better forecasters. TafTrack (Frederick and Amburn, 2000) is such a program. TafTrack is an AWIPS application is provides easily understood information to operational forecasters. Significant information can be obtained from TafTrack contingency tables, average ceiling and visibility category errors, and TEMPO and PROB statistics. Additionally, statistics are available for MOS FWC and MAV guidance which can help forecasters assess their biases and overall value to the TAF. This paper discusses some of the information that can be gleaned from the TafTrack output.

2. MEAN AVERAGE ERROR

The Mean Average Error (MAE) is calculated by comparing the forecast and observed categories of ceiling and visibility. The 24-hour TAF is grouped into four 6-hour periods for this calculation. The resulting MAE by period can be used to identify changes in the TAF quality with time. Comparisons can also be made between the TAF and MOS guidance to determine if/when MOS may generally provide a better forecast. The MAE can also be used to identify and reward forecasters who excel in TAF forecasting. The MAE calculation is:

$$\text{MAE} = (1/n) \sum |(F_i - O_i)| \qquad (1)$$

where n is the total number of forecasts, O_i is an observed category and F_i is the corresponding forecast category.

* *Corresponding author address*: David R. Manning, National Weather Service, 10159 E. 11th St., Suite 300, Tulsa, OK, 74128. e-mail: david.manning@noaa.gov

Guidance forecast categories are used in the TafTrack verification system to provide a means of comparison with the actual TAFs. These categories are shown in Figure 1 for ceiling. The MAV guidance categories were combined to match the less definitive FWC guidance categories.

```
    Visiblity
  #          Range
-------------------------------
  1        0 - 3/8   mile   : MAV cat 1 and 2
  2      1/2 - 7/8   mile   : MAV cat 3
  3        1 - 2 3/4 miles  : MAV cat 4
  4        3 - 5     miles  : MAV cat 5
  5        > 5       miles  : MAV cat 6 and 7

    Ceilings
  #          Range
-------------------------------
  1          < 200    feet
  2      200 - 400    feet
  3      500 - 900    feet
  4     1000 - 3000   feet
  5     3100 - 6500   feet
  6     6600 - 12000  feet
  7     > 12000       feet
```

Figure 1. Ceiling and visibility categories for TAF verification.

Figure 2 shows the overall station MAE for the period from 1 October 2001 through 31 January 2002. The periods, shown on the left, indicate the MAEs for hours 1 through 6, 7 through 12, 13 through 18, and 19 through 24. Labeled across the top are FWC MOS guidance, MAV (Aviation) MOS guidance, and actual TAF. Numbers within the table represent the MAE by category. For example, if ten hours of category 3 were forecast, while category 3 was observed only five times and category 4 was observed five times, the MAE would be 0.50. If category 3 were observed five times and category 5 were observed 5 times, the MAE would be 1.00, i.e., five forecasts of zero error, plus five forecasts of 2 category error, all divided by ten forecasts.

Figure 2 shows that the MAE of the TAF was lower than that of the MAV and FWC in the first 6-hour period (TAF=0.33, MAV=0.42, FWC=0.49). In fact, the TAF had a lower MAE in all four 6-hour periods. Similar MAE statistics can be obtained for visibility, and are shown in Fig. 3.

Absolute Ceiling Errors			
Period	FWC	MAV	TAF
1	0.49	0.42	0.33
2	0.57	0.47	0.45
3	0.60	0.49	0.48
4	0.62	0.53	0.49

Figure 2. Mean absolute errors by category for ceiling at KTUL. FWC and MAV are model guidance. TAF is the official forecast. Periods indicate four 6-hour groups.

Absolute Visibility Errors			
Period	FWC	MAV	TAF
1	0.18	0.12	0.09
2	0.19	0.14	0.12
3	0.20	0.14	0.12
4	0.23	0.13	0.12

Figure 3. Same as Fig. 2 except for visibility errors.

These MAE statistics can be used to help a forecaster (or office) issue a better TAF for the customer. If one of the MOS guidances consistently has a lower MAE than the other MOS or the TAF for a given 6-hour period, the office should begin adjusting their TAFs to more closely match that MOS guidance. It may be that a combination of MOS in the latter hours of the TAF and human forecast in the early hours will result in the best product for the customer. And, because these statistics are available for each forecaster as well as the office, the correct mix of human and MOS input may differ by individual.

Another benefit from these MAE statistics is to recognize and identify the best local TAF forecasters. These forecasters may be able to train or otherwise help others to become better forecasters through seminars, one-on-one training, or through written documentation. It might also be appropriate to recognize the better forecasters through some kind of awards system. To accomplish this, one can simply compare the TAF with the guidance MAE for the interval of interest.

3. CONTINGENCY TABLES

Contingency tables can help forecasters determine biases, both in their own forecasts and also in the MOS forecasts. These tables can help forecasters adjust their focus, and concentrate efforts on ceiling and/or visibility categories that are more challenging. The contingency tables can also indicate those forecast categories in which a forecaster (or MOS) has little or no skill.

Figure 4 is a contingency table of ceiling forecasts for an individual forecaster. TAF ceiling categories are labeled on the left and observed categories are labeled across the top. The total number of forecast hours of a given category are shown on the right side of the table, while the values in the contingency table indicate the percent of time a forecast category was observed. By multiplying the total number of forecasts by a percentage within the table, one can obtain the number of hours for which a particular forecast was valid. In Fig. 4, category 1 was forecast in the TAF four times, was observed 75 percent of that time, or for three hours. Category 2 was observed 25 percent of the time, or for 1 hour.

Ideally, there would be a one-for-one correspondence on the diagonal from the upper left to the lower right. Along that line, a specific category of ceiling would be observed 100 percent of the time when it was forecast. Clearly, this is not the case.

The contingency table in Figure 4 indicates the forecasters had a bias to forecast lower ceilings than were observed, particularly for the lower categories. For example, a category 2 ceiling was observed only 13 percent of the time when forecast, but observed 56 percent of the time as category 3. Category 6 ceilings were observed as category 6 only 21 percent of the time when, but were observed as category 7 in 47 percent of the cases. In fact, numbers (percentages) in the table which fall to the upper right half of the table indicate pessimistic forecasts (forecasts of conditions lower than what were observed). Numbers in the lower left of the table, indicate optimistic forecasts. It would appear the forecasters were being pessimistic with the forecast, preferring to err to the side of safety. However, now that the bias has been identified, forecasters can adjust their thinking process and should be able to make less biased forecasts.

		1	2	3	4	5	6	7	# Fcsts
	1	075	025	000	000	000	000	000	4
	2	005	013	056	012	004	003	008	78
	3	000	017	049	023	006	000	004	262
TAF	4	001	005	018	041	017	006	012	1270
	5	000	001	003	023	028	014	031	824
	6	001	001	001	009	021	021	047	587
	7	000	000	001	003	003	002	090	8759

Figure 4. Contingency table for ceiling forecasts at KTUL. Ceiling categories are indicated across the top and down the left side. Total number of forecasts are indicated on the right. Numbers in the table are percentages.

The contingency table for the Aviation MOS (MAV) is shown in Figure 5. A bias can also be seen in this table, especially for category 2 ceilings. It is interesting to note that for category 5 ceilings, there is a wide range of observed ceilings with relatively high percentages. This would appear to indicate the MAV has limited skill with that forecast category. Forecasters can use this kind of information to alter MOS forecasts, providing better TAFs. Similar information on biases and skill can be observed for visibility from the TAF and MOS forecasts.

		1	2	3	4	5	6	7	# Fcsts
	1	003	021	032	015	000	006	024	34
	2	000	008	059	010	007	000	016	128
	3	000	014	042	023	013	001	007	369
MAV	4	000	004	015	048	016	005	011	777
	5	000	002	005	026	025	009	032	976
	6	000	001	002	007	017	026	048	829
	7	001	000	001	003	003	002	090	8359

Figure 5. Same as Fig. 4, except for the MAV guidance.

4. TEMPO AND PROB STATISTICS

Forecasters have the option of forecasting probabilistic categories in the TAFs. These probabilistic categories include TEMPO and PROB groups. According to the Weather Service Operations Manual (National Weather Service, 1997) TEMPO groups "shall be used to indicate temporary fluctuations to forecast meteorological conditions which are expected to have a high percentage (50 percent or greater) probability of occurrence and last for one hour or less in each instance, and in the aggregate, to cover less than half of the period." PROB groups are used to forecast precipitation events or thunderstorms which have an expected probability of occurrence of 30 or 40 percent.

Figure 6 shows an example of the ceiling verification for these probabilistic forecast groups. The numbers show a strong indication of the proper use of the TEMPO group with 50 percent or greater occurrences for ceiling categories 3, 4, and 7. There was a tendency to over forecast the occurrence of ceiling categories 2, 5, and 6. There were no probabilistic category 1 ceilings were forecast. Visibility statistics show similar trends but are not shown here.

The TEMPO statistics can help forecasters identify the possible overuse of certain categories of ceiling or visibility, particularly at specific TAF sites which may experience local effects. The number of hours that a particular category is observed can also serve as an on-going airport climatology. This type of climatological information can also be helpful in training new forecasters.

Statistics for the PROB40 category show that PROB40 ceilings were not forecast for categories 1, 6, and 7. They were under forecast in categories 3 and 4, and over forecast in category 2. The significant occurrences of lower ceilings indicated in categories 3 and 4, would suggest the TEMPO or prevailing conditions would be more appropriate. The PROB30 category shows better results, with the only category 3 being significantly under. Again, visibility statistics show similar trends.

Since PROB terms are used to forecast stratiform precipitation or convection, the statistics in Fig. 6 will indicate whether the forecasters are properly associating the correct ceiling and visibility categories with the events. For example, Fig. 6 indicates, within the PROB40 group, that 18 hours of category 2 ceiling were forecast, yet there was never an occurrence less than category 3. This may indicate that one or more forecasters has an unrealistic idea as to how low ceilings will become during precipitation events.

	Tempo [Prob50]				Prob40				Prob30				
	.. # of Hours				.. # of Hours				.. # of Hours				
CAT	Fcst	Verif	%	Brier	Fcst	Verif	%	Brier	Fcst	Verif	%	Brier	Ave Brier
1	0	0	-99	9.999	0	0	-99	9.999	0	0	-99	9.999	9.999
2	55	19	35	0.024	18	0	0	0.160	6	2	33	0.001	0.062
3	102	56	55	0.002	77	74	96	0.315	22	14	64	0.113	0.143
4	206	123	60	0.009	93	65	70	0.089	83	24	29	0.000	0.033
5	15	5	33	0.028	3	1	33	0.004	21	5	24	0.004	0.012
6	14	3	21	0.082	0	0	-99	9.999	0	0	-99	9.999	0.082
7	28	20	71	0.046	0	0	-99	9.999	0	0	-99	9.999	0.046
ALL	420	226	54	0.001	191	140	73	0.111	132	45	34	0.002	0.038

Figure 6. Shows data for conditional forecast probability groups from the official forecast at KTUL.

Statistics on visibility may suggest similar findings.

5. SUMMARY

Meaningful TAF verification can generally provide forecasters with information to improve their forecasts to the aviation customer. TafTrack provides this information through the Mean Average Error (MAE), contingency tables, and also through TEMPO and PROB statistics. Managers, such as Science and Operations Officers, can use TafTrack statistics to identify both office and individual weaknesses in TAF issuances which can then be addressed at seminars and in training plans. Information available in the statistics can also be used to develop on-going airport climatologies which can be useful in local studies and also in training new forecasters.

6. REFERENCES

Frederick, J. M., and S. A. Amburn. AWIPS Era Real Time TAF, FWC, and LAMP Ceiling and Visibility Verification Program at WFO Tulsa. Preprints, *Ninth Conf. on Aviation, Range, and Aerospace Meteorology*, Orlando, FL, Amer. Meteor. Soc., 339-344.

National Weather Service, 1997. *Weather Service Operations Manual*, Chapter D-31, Aviation Terminal Forecasts, U.S. Department of Commerce pp. 62.

U.S. Department of Commerce, 1999. *VISION 2005, National Weather Service Strategic Plan for Weather, Water, and Climate Services 2000-2005*, U.S. Department of Commerce, pp. 23.