4.5 TAFTOOLS: DEVELOPMENT OF OBJECTIVE TAF GUIDANCE FOR CANADA AND RESULTS

Jacques Montpetit[1], Pierre Bourgouin[2], Laurie Wilson[1], Richard Verret[2]
[1]Recherche en prévision numérique (RPN), Dorval, Canada
[2]Centre Météorologique Canadien (CMC), Dorval, Canada

1. INTRODUCTION

Aerodrome forecasts (TAFs) are site-specific forecasts that are currently prepared every 6h manually, using guidance from the operational NWP models and the most recent available observations, including radar and satellite data. The Meteorological Service of Canada (MSC) believes that gains in forecast production efficiency can be realised by producing objective TAFs, leaving the final control of the forecast contents with the operational forecaster. Elements of interest to aviation are: ceiling, visibility, weather and obstructions to vision, and wind speed and direction. In the first phase of the project only the two most important elements for aviation, ceiling and visibility, are considered. However, once the tools for those two elements are developed, it will be very easy to extend the development to the remaining elements.

It was decided to use statistical methods, which are relatively easy to develop and run. The statistical software, utilities to produce TAFs from the output of statistical techniques and utilities to edit the resulting TAFs are collectively called TAFTools.

Dallavalle and Dagostaro (1995) have shown that simple persistence is a very strong competitor for very short-range forecasts. Recent work by Vislocky and Fritsch (1997) supports the idea that a system based on observations only should be quite powerful for very-short range forecasting.

Those considerations have led us to attack the TAF forecast problem from two different angles: a component based on observations only for the very-short range forecasts (VSRF) and a Perfect-Prog component based on model output for the short-range forecast (SRF; Montpetit et al., 2002). We expect that the accuracy of the observations-based forecast would deteriorate more rapidly in time than the accuracy of a model-based system.

The presentation will summarise the technique design and show results to date for the very-short range component.

Corresponding author address: Jacques Montpetit, Recherche en prévision numérique (RPN), 2121, route Transcanadienne, Dorval, Qc, Canada, H7P 1J3;
e-mail:jacques.montpetit@ec.gc.ca.

2. THE DATABASE

To prepare a database of observations for development and testing of the statistical relationships, we merged observational data from several sources: 1. Hourly observations (no specials) from the National Climate Data Archive of Canada for the period 1959-1992; 2. Raw observations in ASCII format, both hourlies and specials, for the period 1986-1992; 3. Observations in METAR format from 1993 to the present; and 4. Daily summary observations such as accumulated snowfall and precipitation. The observational database was formed by putting all available observations into a METAR format, then adding the daily summaries. The result was an integrated observation database for 1959 to 1999, of which the first 38 years are used for development and the last three years as an independent test sample.

3. VERY-SHORT RANGE FORECAST

As mentioned previously, the very short-term forecasting technique is based solely on current available observations (Fig. 1). 38 years of hourly observations were used to develop forecast equations relating observations at a time T_0 to observations at a later time T_0+dT (classical statistical formulation) where dT is the forecast projection. We produced forecasts up to 12 hours ahead using that technique. There is one equation for each specific time, that is 24 X 12 equations for each site. As a result, the forecaster will be able to generate a new TAF from any new observation. The predictands consist of all elements of interest to aviation: ceiling, visibility, weather and obstructions to vision, and wind speed and direction. The occurrence of precipitation is a little more tricky to predict objectively because of the large number of possibilities obtained when combining the different types, intensity and convective/non-convective forms. For that reason, we decided to consider 3 separate elements for precipitation : occurrence (yes/no), convection (none, light, moderate, strong), and a reduced set of precipitation types, formed by aggregating some of the observed types into single categories: snow, rain, drizzle, freezing rain and ice pellets, freezing drizzle, rain and snow, thunderstorm.

Predictors include all the weather elements contained in the observation, along with several derived predictors such as 1- and 3-hour tendencies for pressure, temperature, dew-point, and ceiling height. Elements such as 6-hour precipitation accumulation, total 24-hour snowfall, and snow on the ground were also considered. Finally, astronomical factors were included to indicate the day of the year (solar declination) and the time of the day (solar angle).

The VSRF component was developed using each station's specific observations as predictors, with the appropriate time offset as indicated above. Thus a time offset of 2 hours between predictors and predictand provides the forecast equations for a 2-hour forecast. At this stage of development, radar and satellite data have not been integrated into TAFTools yet. Adding a few tendencies as predictors makes up, to some extent, for the lack of data from other stations. The equations were developed using Multiple Discriminant Analysis (MDA).

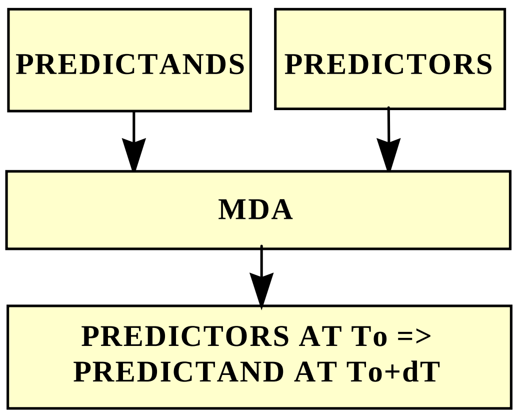

Figure 1. Schematic diagram showing the data flow for the VSRF component of TAFTools.

4. STATISTICAL TOOLS

Aviation operations are based on categories. For example, a ceiling of less than 1000 feet or a visibility less than 3 nautical miles define Instrument Flight Rules (IFR) conditions. Airports have specific operating limits (e.g. a lowest landing limit of 300 feet). A significant impact on aviation operations may result when observed or forecast weather conditions cross different limits. Therefore, aviation forecasts are produced with specific limits, or categories, in mind.

This naturally suggests the use of a statistical method which considers predictands in term of categories rather than in terms of continuous variables. CART and MDA are examples of such techniques. We chose MDA over CART because Burrows and Wilson (2000) showed that MDA gives better results than CART for sky cover, and ceiling is similar to sky cover in the sense that it is a continuous variable categorised for operational purposes. The output of MDA is in probabilistic terms; therefore a post-processing algorithm was added to select a "best" category for prediction. The category selection technique follows the thresholding procedure described by Miller et al. (1979).

The MDA module was developed by RPN. It was validated by comparing its results to those obtained from an independent (commercial) package on a small dataset and also against results from other statistical modules developed at RPN.

We wanted to compare MDA with a standard forecast, so we developed a conditional climatology (CC) technique. The CC technique was developed using categories. The idea is to assign a category to the predictand, then search the database for all occurrences of that same category and record the subsequent evolution of the predictand, still in terms of categories. As an example, suppose that the current ceiling is 800 feet, which is a category 3 in our current setting. The CC module goes through the database, finds all occurrences of ceiling in category 3, and records the ceiling evolution in term of categories over the following 12 hours. That distribution is then converted into probabilities, which are interpreted as a probabilistic forecast of the ceiling, hour by hour.

In order to accelerate the process and diminish the strain on computer resources, the series of probabilities for all predictand categories can be computed only once. New observations are available every day, hour or minute. Therefore, probabilities can be updated once a month to maintain its relevance.

Numerous CC configurations can be tested. We have considered a few of them but the most interesting configuration uses ceiling, visibility, wind direction and 1-hour pressure tendency as predictors (Bourgouin et al., 2002).

Furthermore different parameters can be set (parameter values used in our experiments are indicated within parenthesis):

- observation should be within *n* hours from current time (*n*=0),
- observation should be within *x* days from current date (*x*=40),
- ceiling, visibility, wind direction and speed, and pressure tendency should be within *z* category from initial category (*z*=1).

Finally, two simple forecasting methods were also included: persistence and climatology. All probabilistic results were evaluated using the Rank Probability Score (RPS; Epstein, 1969).

5. RESULTS

Verification with the independent sample showed that ceiling forecasts using CC were initially better than those from MDA using a two-season stratification (Bourgouin et al., 2002) for all projection time (Fig. 3). Verification of visibility forecasts was also in favour of CC (not shown).

Verification by categories showed that MDA was particularly skillful at discriminating the highest (most frequent) category from the others, both for ceiling and visibility. It was decided to exploit that ability by developing a multi-step MDA (MS-MDA). In the first step of MS-MDA, the predictands are re-categorised into two classes, the highest category versus all the other ones combined together.

Equations were re-developed and forecasts produced using that set-up. In the second step, another MDA was developed on the lowest categories only using a database excluding the highest category observations. Forecast from both MDAs were then combined. RPS from MS-MDA improved markedly over regular MDA. MS-MDA scores were slightly superior to those from CC for ceiling forecast and very close for visibility (Fig. 4 and 5).

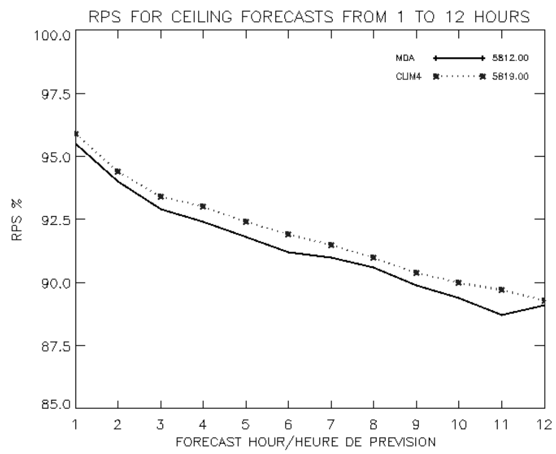

Figure 3. Rank probability scores (RPS) for probabilistic ceiling forecasts using CC (dotted line) and MDA (full line) for projection time 1 to 12 hours based on 2200 UTC data. Scores are averaged for 8 Canadian sites obtained from a 3-year independent sample.

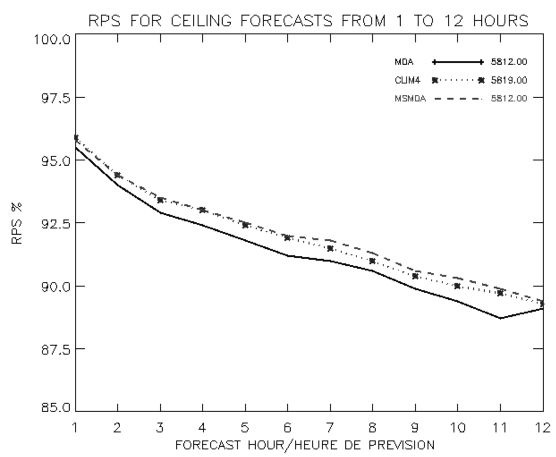

Figure 4. As in Fig. 3, but comparing 3 techniques: MDA using a 2-season data stratification (MDA), MS-MDA (MSMDA), and CC using ceiling, visibility, wind direction and pressure tendencies as predictors (CLIM4).

We also compared results from MS-MDA and CC to those from persistence and climatology. MDA produced much better forecasts than persistence (Fig. 5). As suggested above, this is encouraging since persistence is considered difficult to beat during the first few hours.

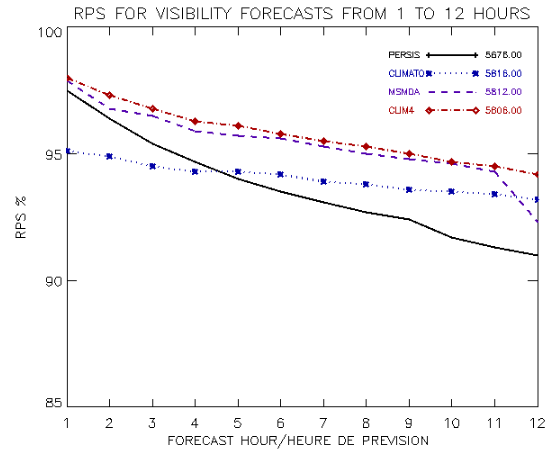

Figure 5. As in Fig. 4, but for probabilistic visibility forecasts comparing four different techniques: simple persistence (PERSIS), climatology (CLIMATO), CC (CLIM4), and MS-MDA (MSMDA).

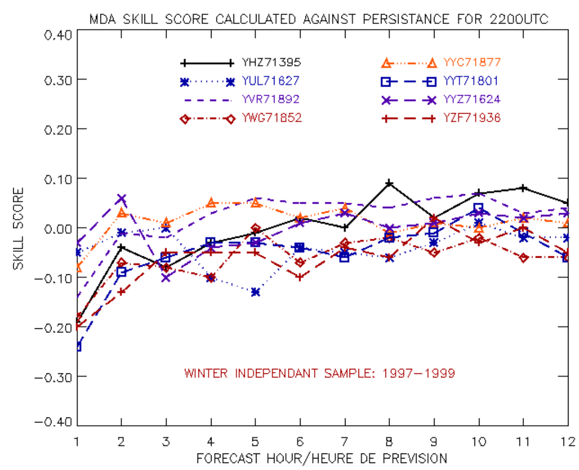

Figure 6. Heidke skill scores of ceiling categorical forecasts based on 2200 UTC data. Scores have been calculated individually for each station. The standard for comparison is persistence.

A selection procedure was also developed to provide forecasters and the aviation community with a categorical forecast. It is based on the unit bias model described by Miller et al. (1979) and was developed with the same development database used for the statistical models. Categorical forecasts are evaluated using persistence as the standard of comparison. Even though MDA probabilistic forecasts are superior to persistence at all projection time (Fig. 5), the thresholding method is unable to translate this characteristic to the selection procedure (Fig. 6).

This result could be better understood bearing in mind the following points: First, a day-to-day subjective evaluation clearly shows that MDA probabilistic forecasts contain more information than the categorical ones. The category selection procedure is simplistic and results in a significant loss of information compared to the probabilistic forecasts. It is believed that a more sophisticated selection procedure would produce better categorical forecasts than the ones from persistence. Second, the Heidke Skill score does not give points for being close to the observed category nor does it value the forecast of trends. A subjective evaluation indicates that MDA has skill at predicting trends, even if the resulting forecast category was not the exact observed one. Finally, the number of categories was chosen to meet the forecaster operational requirements. No analysis was performed to determine a optimal number of categories. We now believe that fewer categories would improve the system.

6. CONCLUSION

In an effort to improve production efficiency of terminal aviation forecasts in Canada, a project was undertaken to produce objective TAFs. The project relies on statistical methods to produce a very-short range forecast from observations only.

The statistical method chosen for the VSRF module was Multiple Discriminant Analysis (MDA). That method is well suited to the specific problem of aviation forecasting. A conditional climatology method was also developed. It was felt that this approach should also produce good results and it would be useful as a benchmark. Verifications showed that CC was indeed a strong competitor to MDA.

Based on the fact that MDA was skillful at discriminating the large-sized categories from the small-sized, it was decided to test a multi-step procedure, MS-MDA. Results from a two-step MDA showed a significant improvement over regular MDA because the former enhances the signal from small-sized categories by separating them from large-sized categories.

Interestingly, MS-MDA does not use the same set of predictors at each step. For example, ceiling height is the most frequent predictor at step one while it is never used at step two. In fact, MS-MDA recognises that different physical parameters are needed, for example, to distinguish the existence of a ceiling compared to those needed to determine its height.

Since CC shows higher RPS values than MDA, one may ask why not simply use CC. The main reason is that a subjective evaluation of verifications by category suggests that MDA is slightly sharper than CC, a desirable attribute of the forecasts which would not be revealed by the RPS results.

Finally, both CC and MS-MDA produce much better probabilistic forecasts than simple climatology or persistence. This makes these two techniques valuable for short-range forecasting.

Currently, efforts are directed at the selection procedure to improve the extraction of information for the improvement of categorical forecasts.

8. REFERENCES

Bourgouin, P., J. Montpetit, L. J. Wilson and R. Verret, 2002: TAFTOOLS: DEVELOPMENT OF OBJECTIVE TAF GUIDANCE AT CMC, Part one - Introduction and development of the very short-range forecast module. Preprints, *16th Conf. On Probability and Statistics In the Atmospheric Sciences*, Orlando, FL, Amer. Meteor. Soc., 94-98.

Burrows, W.R., and L.J. Wilson, 2000: Comparison of CART and MDA for statistical sky cover forecast models. Preprints, *15th Conf. on Probability and Statistics in the Atmospheric Sciences*. Asheville, NC, Amer. Meteor. Soc., 40-43.

Dallavalle, J.P., and V.J. Dagostaro, 1995: The accuracy of ceiling and visibility forecasts produced by the National Weather Service. Preprints, *Sixth Conf. On Aviation Weather Systems*, Dallas, TX, Amer. Meteror. Soc., 213-218.

Epstein, E.S., 1969. A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor*., **8**, 985-987.

Miller, R. G., and D. L. Best, 1979: A model for converting probability forecasts to categorical forecasts. Preprints, *Sixth Conference on Probability and Statistics in Atmospheric Sciences*, Banff, Alberta, Canada, Amer. Meteor. Soc., 98-102.

Montpetit, J., P. Bourgouin, L. J. Wilson and R. Verret, 2002: TAFTOOLS: DEVELOPMENT OF OBJECTIVE TAF GUIDANCE FOR CANADA, Part two - DEVELOPMENT OF THE SHORT-RANGE FORECAST MODULE AND RESULTS . Preprints, *16th Conf. On Probability and Statistics In the Atmospheric Sciences*, Orlando, FL, Amer. Meteor. Soc., 99-101.

Vislocky, R.L., and J.M. Fritsch, 1997: An automated, observations-based system for short-term prediction of ceiling and visibility. *Wea. Forecasting*, **12**, 31-43.