TAFTOOLS: DEVELOPMENT OF OBJECTIVE TAF GUIDANCE FOR CANADA

Part one - Introduction and development of the very short-range module

Pierre Bourgouin[1], Jacques Montpetit[2], Richard Verret[1], Laurie Wilson[2]

[1]Centre Météorologique Canadien (CMC), Dorval, Canada

[2]Recherche en prévision numérique (RPN), Dorval, Canada

1. INTRODUCTION

Aviation terminal 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 local 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, will be 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 to run. The statistical software, the utilities to produce TAFs from the output of statistical techniques and the 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 (Fig. 1): 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). We would expect that the accuracy of the observations-based forecast would deteriorate more rapidly in time than the accuracy of a model-based system.

In addition to the VSRF and SRF systems, a third component will blend the two systems together to produce a single TAF. Results from independent sample tests will indicate the appropriate weights to place on the output of the VSRF and SRF forecasts at each projection. The presentation will summarise the technique design and show results to date for the very-short range component. For details on the short-range module see Montpetit et al. (2002).

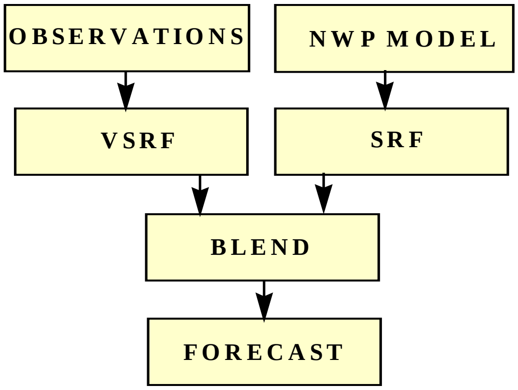

Figure 1. Schematic flow diagram showing the three main components of TAFTools, VSRF, SRF and BLEND.

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 2000, of which the first 38 years is used for development and the last three years is used 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. 2). The 38 years of

---

Corresponding author address: Pierre Bourgouin, Centre Météorologique Canadien (CMC), 2121, route Transcanadienne, Dorval, Qc, Canada, H7P 1J3;
e-mail:pierre.bourgouin@ec.gc.ca.

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 only the data for the predictand station as predictors, with the time offset as indicated above. Thus a time offset of 2 hours between predictors and predictand gives forecast equations for a two hour forecast. A forecaster can look at radar or satellite images, but our system will not use such data initially. Adding a few tendencies as predictors to some extent makes up for the lack of data from other stations. The equations were developed using Multiple Discriminant Analysis (MDA).

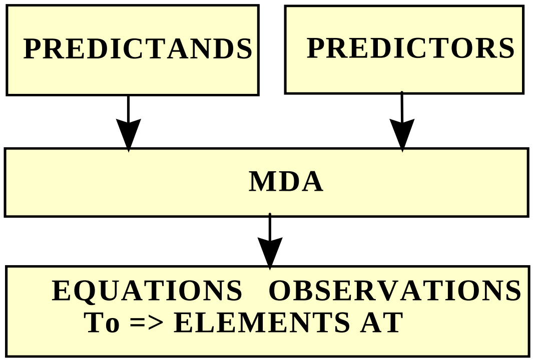

Figure 2. Schematic diagram showing the data flow for the VSRF component of TAFTools.

4. BLENDING

Comparisons between scores obtained from the VSRF and SRF systems on the independent sample will indicate relative weights to give to each components as a function of time. We expect that scores will depend on the specific location, season and element considered (e.g. ceiling or visibility).

There is another important factor to consider: the issue time. TAFs are generally issued around 0430, 1030, 1630 and 2230 UTC. On the other hand, NWP models are generally run every 12 hours, usually shortly after 1200 and 0000 UTC. In designing the blending equations, we will take into account the ability to use the VSRF component, which can be initialised on the latest hourly data, as an update to the (usually) less current model-based forecasts. For example, NWP data for the 2230 UTC TAF issue time are 10 hours old. In that case, a 7-hour forecast from the VSRF component has to be compared with a 17-hour forecast from the SRF component.

The type of transition from the TAF produced with the VSRF component to the SRF-TAF, as expressed by the relative weights, remains to be decided: It can be abrupt, progressive or favour the worse case scenario. However, we also want to retain some flexibility in the sense that the forecaster should be able to select the TAF associated with the VSRF (SRF) component for a longer (shorter) term if the forecaster feels that it is a better solution for a specific situation.

In the future, the blending component will become more than a mere merging between the VSRF and SRF components. We expect to add more sources of information : lightning detector, satellite, radar, direct model output, etc. Also, we will consider grouping stations and adding local forecasting rules (expert system).

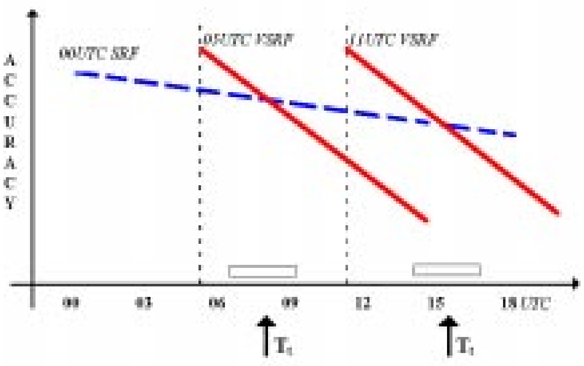

Figure 3. Schematic diagram showing the gradual decline in accuracy of the 00UTC SRF (dashed line) and the 05 and 11UTC VSRF (continuous lines) as a function of time. Two different transition times (T_t) are indicated for two TAF issue times (05 and 11 UTC). The two boxes represent transition windows. Note that the SRF module is initialised once while the VSRF module is initialised twice.

5. 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 of the results of tests (Burrows and Wilson, 2000) which 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 that is categorized for operational reasons. The output of MDA is in probabilistic terms; we will add a post-processing algorithm which will select a “best” category for prediction. The category selection technique is currently under development; we are following the thresholding procedures described by Miller (1981).

The MDA module was developed by RPN. It was validated by comparing its results to those obtained from a 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 data-base, 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.

Of course, it would be easy to produce the distributions for all the different possibilities. This would accelerate the process and be much less demanding in computer resources. We have focused on the ceiling and visibility, which are the most important elements for a TAF.

Many different CC configurations can be tested. We have considered a few of them, which are described below. If there is more than one predictor, all predictors must be found in their respective category for a particular observation to be selected.

- clim1 : uses the predictand as the only predictor i.e. ceiling (visibility) is used to forecast the ceiling (visibility).
- clim2 : uses ceiling and visibility to forecast either ceiling or visibility.
- clim3 : uses the predictand and wind direction to forecast ceiling or visibility.
- clim4 : uses ceiling, visibility, wind direction, one hour pressure tendency to forecast either ceiling or visibility.

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).

Comparison among the 4 configurations showed that clim3 and clim4 give the best results and that clim4 has a slight edge over clim3. For that reason clim4 is used as our standard of comparison. Finally, two simple forecasting methods were also included: persistence and climatology. All probabilistic results were evaluated using the Rank Probability Score (RPS, Epstein 1969).

6. RESULTS

The CC gave impressive results for ceiling and visibility, especially the clim4 version. The RPS values from clim4 were initially better than those from MDA (Fig. 4). It was thought that part of the difference was related to a difference in the data stratification used the develop the equations. MDA initially considered a 2 season data (summer, winter) stratification while CC used data within a moving 80-day window centered on current day. MDA was redone using a moving window of 3 months (previous, current and following month). Unfortunately, RPS scores did not improve much and were still inferior to those from CC (Fig. 4).

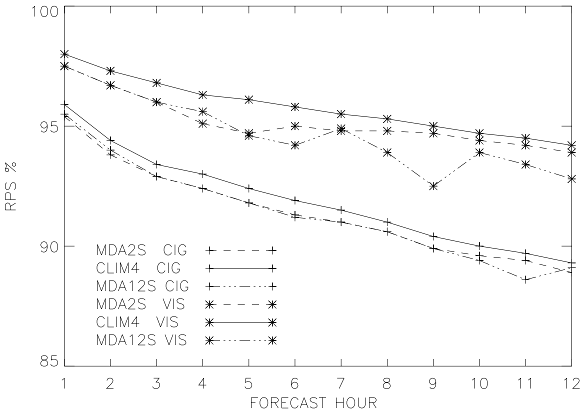

Figure 4: Rank probability scores (RPS) for categorical ceiling (+) and visibility (*) forecasts, for forecasts from 1 to 12 hours based on the 2200 UTC data. Scores are averaged for 8 Canadian sites obtained from a 2-year independent sample. Three different techniques are compared: CC using the clim4 set-up (CLIM4), MDA using a 2-season data stratification (MDA2S), and MDA using a 12-season data stratification (MDA12S).

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. 5).

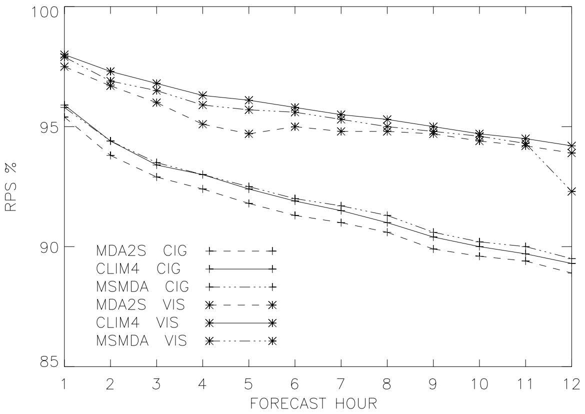

Figure 5: As in Fig. 4, but comparing 3 techniques: MDA using a 2-season data stratification (MDA2S), MS-MDA (MSMDA), and CC using the clim4 set-up (CLIM4).

We also compared results from MS-MDA and CC to persistence and climatology. The former produced much better forecasts than the latter (Fig. 6). As suggested above, this is encouraging since persistence is considered as difficult forecast to beat during the first few hours.

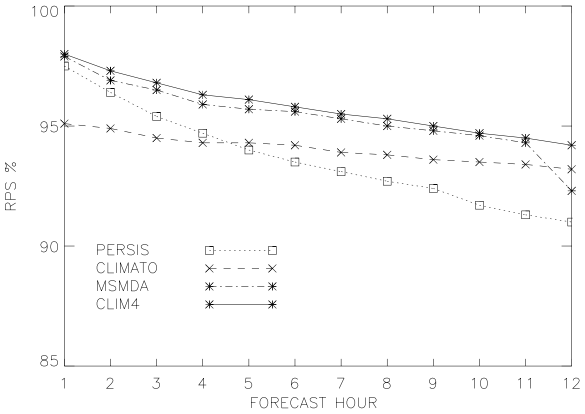

Figure 6: As in Fig. 4, but for categorical visibility forecasts comparing four different techniques: simple persistence (PERSIS), climatology (CLIMATO), CC using the clim4 set-up (CLIM4), and MS-MDA (MSMDA).

7. 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 two separate forecasts, one aimed at the very-short range from observations only and the other for the short-range from NWP model output.

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. In order to improve results obtained from MDA, data stratification was changed from 2 seasons to a moving window of 3 months (previous, current and following month). This modification was not conclusive.

Based on the fact that MDA was skillful at discriminating the highest categories from the lowest, it was decided to test a multi-step procedure, MS-MDA. Results from a 2-step MS-MDA showed a significant improvement over our regular MDA. To understand that improvement, consider the following for ceilings (the same holds true for visibilities). The regular MDA answer just one question: what is the ceiling height? The MS-MDA discriminates between the highest category and the others. The highest category is basically the “no-ceiling” category. Therefore, the MS-MDA answers two questions. In its first step, is there a ceiling? And in the second step: what is its height? MS-MDA does not use the same set of predictors to answer the two questions. For example, the ceiling height, a derived element, is the most frequent predictor to answer the first question while it is never used to answer the second one. In fact, MS-MDA recognises that different physical parameters are at work for the two different questions.

Since CC shows higher RPS values than MDA, one may ask why not simply use CC. The answer is two fold. First, in an operational setting, MDA is much easier to use since only a series of coefficients are necessary to produce the forecasts. CC requires exploration of a large database for the preparation of each forecast, which requires significant space and time. Second, 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 forecasts than simple climatology or persistence. This makes these two techniques valuable for short-range forecasting.

8. REFERENCES

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. Meteror. 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., 1981: GEM: A Statistical weather forecasting procedure. *NOAA Technical Report NWS 28*. National Oceanic and Atmospheric Administration, U.S. Department of Commerce, November 1981, 103pp.

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, American Meteorological Society.

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.