A Fuzzy Logic–Based Analog Forecasting System for Ceiling and Visibility

BJARNE HANSEN

Cloud Physics and Severe Weather Research Section, Meteorological Research Division, Environment Canada, Dorval, Quebec, Canada

ABSTRACT

WIND-3 is an application for aviation weather forecasting that uses the analog method to produce deterministic predictions of cloud ceiling height and horizontal visibility at airports. For data, it uses historical and current airport observations [routine aviation weather reports (METARs)], and model-based guidance. It uses the perfect prognosis assumption as it is designed to use any model-based predictions of wind direction and speed, temperature and humidity, and precipitation occurrence and type to specify conditions in the 1–24-h projection period. To identify and rank analogs, according to their degree of similarity with the present situation, it uses a fuzzy logic–based algorithm to measure similarity between past situations, which are complete series of METARs, and current situations, which are a composite of recent METARs and model-based guidance. It uses the retrieved analog ensemble, the set of most similar analogs, to make predictions of ceiling and visibility in the 1–24-h projection period. WIND-3 has been tested by being run continuously in real time for 1 yr, producing forecasts for 190 major Canadian airports. It produces accurate forecasts, based on summaries of Heidke skill score (HSS) statistics, and compared to two benchmarks, persistence and official aerodrome forecasts [terminal aerodrome forecasts (TAFs)]. WIND-3 predictions of instrument flight regulation (IFR) conditions in the 0–6-h period have an HSS of 0.56, and in the 7–24-h period have an HSS of about 0.40, compared to respective HSS scores for persistence forecasts of 0.53 and less than 0.20.

1. Introduction

The safety and efficiency of air travel depends on accurate and timely forecasts of airport weather. Pilots use these forecasts when deciding how much reserve fuel to load on board before takeoff. If poor weather is forecast at a flight's destination airport, then there is an increased chance that the flight will have to be diverted en route to an alternate airport. In such cases, pilots will load on extra fuel to extend the flight range of the airplane to reach these alternate airports.

The Meteorological Service of Canada (MSC) is responsible for providing accurate and timely weather forecasts for 190 airports across Canada. The forecasts describe weather conditions expected to affect flight conditions for up to the next 24 h. These conditions include cloud ceiling height, horizontal visibility, pre-cipitation, and wind direction and speed. Forecasters work to keep these forecasts as accurate and current as possible, and will quickly revise forecasts as appropriate for the current weather situation using all relevant information. The most frequent cause for forecasts needing to be revised is an unanticipated change in ceiling or visibility (H. Stanski 1999, personal communication).

At every airport in Canada, large archives of past weather observations (of variables listed in Table 1), often stretching back several decades, contain climatological information on ceiling and visibility specific to the airport. Forecasts for the present case can be based on similar past cases using the analog method. To take advantage of this climatological data and the analog method, a tool named WIND-3 has been developed that quickly finds, summarizes, and displays the ceiling and visibility values of past weather cases that are most similar to the present case, i.e. analogs. The display provides forecasters with a new form of relevant information that helps them to improve forecast accuracy.

The work described here builds upon that of Hansen (2000) and Riordan and Hansen (2002), in which it was

Corresponding author address: Bjarne K. Hansen, Cloud Physics and Severe Weather Research Section, Meteorological Research Division, Environment Canada, 2121 Trans-Canada Highway, Dorval QC H9P 1J3, Canada.
E-mail: bjarne.hansen@ec.gc.ca

TABLE 1. Aerodrome observational data. Regular hourly observations from airports (METARs) contain six types of data described by 12 attributes.

Type	Attribute	Units
Temporal	Date	Julian date of year (wraps around)
	Hour	Time UTC
Ceiling and visibility	Cloud amount(s)	Tenths of cloud cover (for each layer)
	Cloud ceiling height	Height in meters of $\geq$6/10ths cloud cover
	Visibility	Horizontal visibility in meters
Wind	Wind direction	Degrees from true north
	Wind speed	kt
Precipitation	Precipitation type	None, rain, snow, etc.
	Precipitation intensity	None, light, moderate, heavy
Dewpoint spread and temperature	Dewpoint temperature	°C
	Dry-bulb temperature	°C
Pressure	Pressure trend	kPa h^{-1}

FIG. 1. Locations of 190 airports for which MSC regularly issues TAFs.

shown that the analog method, implemented with a fuzzy k-nearest neighbors (k-NN) algorithm, and supplied with large database of in situ weather observations [routine aviation weather reports (METARs)], can skillfully predict cloud ceiling and visibility at an airport. Subsequently, the analog method has been adapted into a system to provide ceiling and visibility forecast guidance in real time for all major airports in Canada.

The rest of this section introduces the problem of the objective forecasting of ceiling and visibility, and describes some previously tried approaches. The second section describes a new approach that combines the analog method with fuzzy logic. The third section reports the results. The final section lists conclusions and recommended subjects for future work.

a. Ceiling and visibility forecasts

The Meteorological Service of Canada (2006) defines a terminal aerodrome forecast (TAF) as follows: "the forecaster's best judgment of the most probable weather conditions expected to occur at an aerodrome together with their most probable time of occurrence. It is designed to meet the preflight and in-flight requirements of flight operations. Aerodrome forecasts are intended to relate to weather conditions for flight opera-
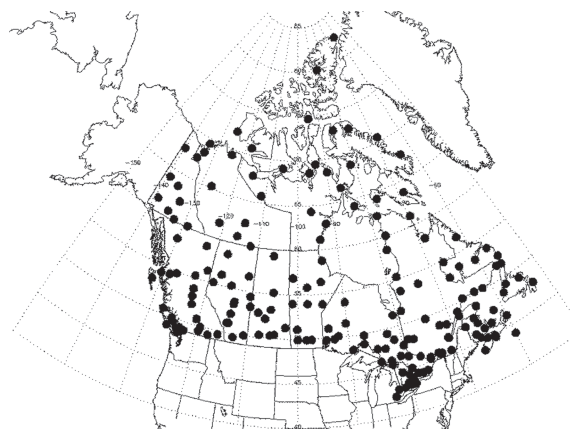
tions within 5 nautical miles of the centre of the runway complex depending on local terrain."

TAFs are regularly produced and issued for 190 airports in Canada (Fig. 1). Forecast timings of significant changes in flight conditions are supposed to be accurate to within 1 h. Cloud ceiling height and horizontal visibility (defined by Glickman 2000), hereafter referred to simply as ceiling and visibility, are the two variables that together determine flight category [e.g., instrument flight rules (IFR) or visual flight rules (VFR), defined below in section 3]. Forecast flight categories at airports are used by airlines to make operational fueling decisions and are used by air traffic managers to anticipate airport capacities (i.e., manageable rates of airplane acceptances and departures).

In addition to forecasts of ceiling and visibility, TAFs include information about precipitation occurrence and type, and wind direction and speed. TAFs must be quickly revised when observed conditions differ significantly from forecast conditions. TAFs are challenging to produce due to the expected precision: when poor flight conditions are forecast, ceiling height forecasts are expected to be accurate to within 100 ft and visibility forecasts are expected to be accurate to within ¼ mi; and when significant changes of flight category are forecast, the forecast time of the change is expected to be accurate to within 1 h.

The goal of the research described here is to develop a useful tool for operational forecasters to help them in predicting ceiling and visibility more efficiently and more accurately.

b. Direct model output

Operational forecasters have a number of objective techniques to help them to forecast ceiling and visibility. Most of these techniques combine the respective

strengths of two basic, complementary forecast techniques: dynamical and statistical (e.g., Jacobs et al. 2003). Dynamical techniques (numerical models) suggest probable ambient conditions related to the ceiling and visibility over large-scale areas, whereas statistical techniques using observations made at any specific airport suggest probable values of the ceiling and visibility themselves in small-scale areas around the specific airport. Current numerical models are very useful for predicting conditions related to ceiling and visibility, but are not as useful for predicting the values of ceiling and visibility themselves for the following two reasons.

First, large-scale numerical models do not factor in or reflect the particular details of the subgrid-scale local airport climatology. For example, a comparison of annual frequency of fog (hours of visibility of ≤½ mi in fog) for several neighboring airports in Canada shows that frequency can vary by a factor of 2 across a distance of one grid-scale unit (15 km), particularly near coasts.

Second, dynamical model resolution, in the horizontal and vertical, is currently inadequate for modeling cloud ceiling height and horizontal visibility. For example, in the Global Environmental Multiscale (GEM) model, there are approximately only three model levels below 1000-ft altitude (MetEd 2006). When ceiling heights are forecast below 1000 ft, the goal is to forecast them accurately to within 100 ft. The horizontal scale of interest for aviation forecasting is 5 n mi (8 km) around the aerodrome, which is approximately half that of the current 15-km GEM grid scale (MetEd 2006). Variables of primary concern to aviation interests are particularly challenging to forecast with models alone, namely very short-range precipitation rate and type, cloud structure, and fog (Stoelinga and Warner 1999). Models' basic variables are wind vectors (u and v), humidity, geopotential height of eta levels, and surface pressure (to fix the surface boundary). Highly reliable values of weather elements that affect ceiling and visibility may be obtained directly from model output (e.g., wind speed and direction, temperature, and humidity), and fairly reliable forecasts of precipitation occurrence and type are obtainable from models, to help in forecasting the ceiling and visibility. However, the ceiling and visibility themselves may only be indirectly inferred from direct model output because for the same ambient conditions (observed weather other than ceiling and visibility) wide distributions of ceiling and visibility are often observed. All subgrid-scale physical processes are parameterized, particularly those involving atmospheric water, which is the key factor to consider in forecasting the ceiling and visibility.
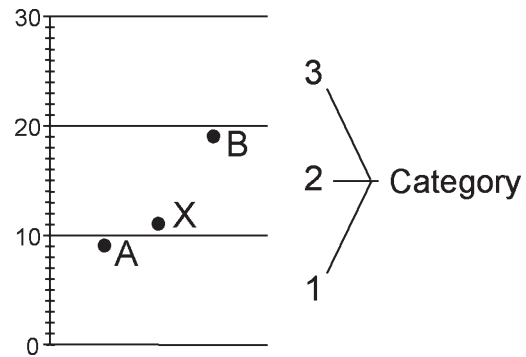

FIG. 2. Crisp (nonfuzzy) categories tend to misrepresent data and thus systematically lose information about the degree of similarity between compared data items.

c. Conditional climatology

A climatological forecast is defined as one "based solely upon the climatological statistics for a region rather than the dynamical implications of the current conditions" (Glickman 2000). The climatological forecast method has been applied to develop numerous systems for generating objective guidance for forecasting ceiling and visibility (Martin 1972; Stutchbury and Hawkes 1974; Lund and Grantham 1979; Lund and Tsipouras 1982; Whiffen 1993; M. A. Purves 1997, personal communication).

Previous implementations of the conditional climatology lack relevance and specificity for specific weather situations due to two factors. First, the previous systems were based on summaries of preselected categories, crisply defined broad categories, and may therefore fit poorly with the particulars of the current weather situation. Basically, continuous numbers are better than coarse categories for describing variables. For example, referring to Fig. 2, if observation X represents the current situation, and observations A and B are potential past analogs, then a continuous-number measure will correctly determine that A is four times closer to X than is B, whereas a coarse category description would misleadingly group B in the same category as X, and the closer A in a separate category from X. Most conditional climatology systems are based on the use of such coarse categories (e.g., four seasons with sharp thresholds). In contrast, WIND-3 uses continuous fuzzy sets to evaluate a degree of similarity between compared observations (see Fig. 3). For each hour in the 0–24 h period, all potential analogs are evaluated according to their degree of similarity with the observed or expected conditions. The most similar analogs are identified and ordered according to a decreasing degree of similarity, and form an *analog ensemble*, a group of past observations with conditions most similar
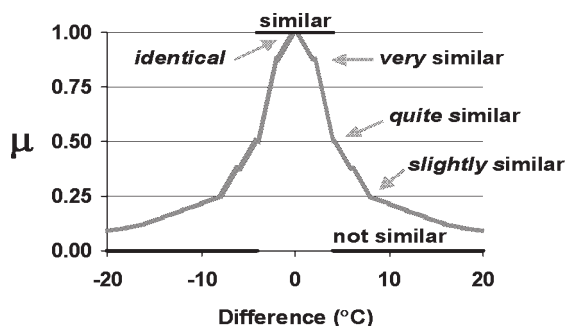
FIG. 3. A fuzzy set for measuring the degree of similarity, $\mu(x)$, between two variables. The greater the difference between two variables, x, the lower the value of the degree of similarity. The parameter μ is continuous and can assume values in the range $0.0 \leq \mu \leq 1.0$, where particular values can correspond to quantitative descriptions of similarity, such as *very*, *quite*, and *slightly*. It contrasts with a discontinuous binary function (similar or not similar).

to current and forecast conditions, from which expectations about the ceiling and visibility may be based. This is further explained below.

Second, previous systems do not integrate specific data from current actual observations and valid model guidance. In contrast, WIND-3, in real time upon the receipt of each new airport observation, integrates the observation with specific conditions describing the current weather case, including weather variables forecast for the 1–24 h period from models, and ambient conditions related to the ceiling and visibility (precipitation, wind, humidity, and temperature). The results are available for forecasters through a Web browser in textual and graphical summaries within seconds of the receipt of the latest observation. This increases the convenience of the conditional climatology and thereby promotes the use of the technique. This is further explained below.

Hastie et al. (2001) distinguish between two basic methods for statistical learning and prediction: 1) parametric linear models and least squares, and 2) nonparametric *k*-NN methods. Based on a review of postprocessing systems for prediction of the ceiling and visibility, the former approach has been applied and tested much more than has the latter (e.g., Glahn and Lowry 1972; Bocchieri and Glahn 1972; Bocchieri et al. 1974; Wilson and Sarrazin 1989; Vislocky and Fritsch 1997; Leyton and Fritsch 2003; Leyton and Fritsch 2004; Jacobs and Maat 2005). The systems reviewed are largely (but not completely) based on the former type of method. In the review, no systems could be found based on the later type of method. WIND-3 applies and tests the nonparametric *k*-NN method, which in meteorological terms may be thought of as a basic analog method.

2. Method

The previous section introduces the problem of objective forecasting of the ceiling and visibility, and describes some previously tried approaches. This section describes a new approach to the problem that combines two previously untried methods, analog forecasting and fuzzy logic, in the WIND-3 system.

a. Analog forecasting and fuzzy logic

Analog forecasting is defined as a "method of forecasting that involves searching historical meteorological records for previous events or flow patterns similar to the current situation, then making a prediction based on those past events or patterns" (Glickman 2000). Because *similarity* can be perceived and defined in many different ways, forecasting systems that use the analog method are diverse. The analog method appeals intuitively to forecasters who, when faced with a seemingly familiar forecast situation, wonder how similar situations evolved in the past.

An advantage of the analog method, compared to other methods, is that when its results appear doubtful, they are relatively easy to trace and reconstruct. The observations that solutions are based upon can be kept throughout the processing routine and can be available for inspection along with the summary solutions. If a forecaster doubts an analog forecast of a ceiling, he or she can inspect the particular analogs used to make the forecast, reject some of them for being poor analogs for the current situation, and revise the solution accordingly.

Analog forecasting can be achieved through the application of a *k*-NN algorithm. The most similar analogs for a given weather situation, the *k*-NN, may be treated as an *analog ensemble*. The analog method applied in WIND-3 is essentially an application of *k*-NN, which is described by the online encyclopedia Wikipedia (cited 2006; available online at http://en.wikipedia.org/wiki/K-nearest_neighbor_algorithm) as follows:

In pattern recognition, the *k*-nearest neighbor algorithm (*k*-NN) is a method for classifying objects based on closest training examples in the feature space. The training examples are mapped into multidimensional feature space. The space is partitioned into regions by class labels of the training samples. A point in the space is assigned to the class *c* if it is the most frequent class label among the *k* nearest training samples. Usually Euclidean distance is used. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the actual classification phase, the same features as before are computed for the test sample (whose class is not

known). Distances from the new vector to all stored vectors are computed and k closest samples are selected. The new point is predicted to belong to the most numerous class within the set.

With WIND-3, "class" refers to values or ranges of values of the ceiling and visibility in k-NN. For example, given two classes, IFR and VFR, if the majority of a case's nearest neighbors are IFR, then the case itself can be predicted to be IFR. How the distance measure is developed and applied is described in detail below in section 3c.

In describing the k-NN method, Hastie et al. (2001) explain that whereas "the linear model makes huge assumptions about structure and yields stable but possibly inaccurate predictions [the] method of [k-NN] makes very mild structure assumptions: its predictions are often accurate but can be unstable."

Two important issues in the application of the k-NN method are the selection of an optimal value of k and the specific choice of the distance measure. The instability of solutions from the k-NN technique increases as k decreases: when k equals 1, the nature of every point being modeled is determined by that of its single nearest neighbor in the training data, and slight changes in independent variables often change the selection of the nearest neighbor and thus the solution, whereas when k is higher, the nature of every point being modeled may be determined by that of the majority of its k nearest neighbors, and slight changes in independent variables seldom change the solution. In preliminary experiments, using a database of approximately 300 000 hourly airport observations, by varying the value of k, it was found that a value of $k = 16$, as a default setting, gave the best results in terms of summary statistics; or in other words, an analog ensemble that consists of the 0.005% of past observations most similar to the current situation gives generally best results for k-NN (Hansen 2000, section 4.2). For testing WIND-3 with a large number of forecasts, $k = 16$ was applied, but for any individual forecast, one could specify any value for k.

Although Euclidean distance measures are usually applied for k-NN, a fuzzy logic–based similarity measure was applied for WIND-3 because the analog method depends on an evaluation of the overall similarity based on a comparison of numerous, distinct continuous and categorical properties, and fuzzy logic has proven particularly effective for such problems in a wide variety of domains (Hansen 2000, section 2.4). WIND-3 uses fuzzy logic to emulate an expert assessing the *degree of similarity* between current and past weather situations, where the degree of similarity is represented by a value that ranges continuously along a scale from 0.0 (completely dissimilar) to 1.0 (identical),

with relative degrees of similarity describable by numerical values anywhere along the scale, and by corresponding fuzzy words. For example, referring to Fig. 3, if two temperatures differ by 2°C, then $\mu = 0.9$; if they vary by 4°C, then $\mu = 0.5$; and if they vary by 8°C, then $\mu = 0.25$. In these operations, μ represents a similarity index, the values of which correspond to respective qualitative degrees of similarity: *very*, *quite*, and *slightly*. The fuzzy set contrasts with the "crisp set" drawn as a discontinuous function (Figs. 2 and 3), which is essentially what is applied in all previous conditional climatology systems for the ceiling and visibility, and which systematically loses information about the degree of similarity between compared data items, information that is of value for analog forecasting (e.g., the weight of any analog's contribution to a forecast can be weighted according to its degree of similarity to the current case). For a theoretical basis of a fuzzy k-NN algorithm, the reader can refer to Keller et al. (1985). WIND-3 consists of three parts that are described in the following subsections: data, a fuzzy similarity-measuring algorithm, and a forecast composition step.

b. Data

The following data for 190 aerodrome sites are used:

1) Archives of aerodrome observations (METARs) describing past weather conditions. Only regular hourly observations are currently readily available and used ("special" observations are not used). All the routinely observed variables listed in Table 1 are available. The data were subjected to thorough quality control and pertain to the period from 1971 to 2004 inclusively. Records for major airports are more than 99% complete. Records for most sites are more than 90% complete. Only 20 airports have records less than 50% complete, and these are generally minor airports whose observation programs began during the later part of the period. The data is described in the National Climate Data and Information Archive (2006).

2) Recent aerodrome observations describing recent and present weather conditions. All observations are available, both regular hourly observations and additional special observations, which are made whenever conditions change significantly. For regular hourly observations, all of the routinely observed variables listed in Table 1 are available. The data used to compose the present case include the latest two regular hourly observations plus any special observations received since the latest whole hour.

3) GEM model–based, MOS-based forecast guidance describing future conditions. These data are valid
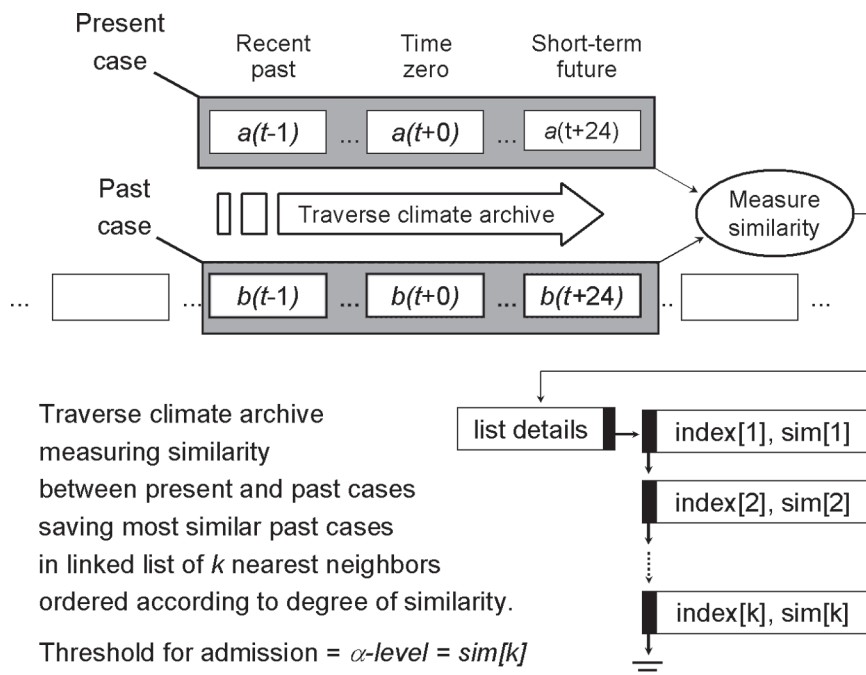
FIG. 4. Analog retrieval is achieved by measuring the similarity between the present composed case and all past cases. All comparable variables of cases are compared, pair by pair, and the lowest similarity value describes the overall similarity between the two cases, sim.

for the 0–24-h projection period. Forecast values of all the routinely observed variables listed in Table 1 are available and used, except for the ceiling and visibility themselves. The model-based forecast variables help WIND-3 to anticipate changes in conditions that often signal changes in ceiling and visibility, such as wind shifts, or the onset or cessation of precipitation.

The reception of each new airport observation triggers the composition of a current "case" and the production of an analog forecast. What is known from a new observation (for time $t - 0$ h) and what is inferred from model-based guidance (for $t + 1$ h, ..., $t + 24$ h) is combined as a "present case" (Fig. 4). Basically, in the 0–6-h time frame, conditions are interpolated from completely observation-based at $t - 0$ h to completely model-based at $t + 6$ h (and from there on to $t + 24$ h).

The GEM model–based, MOS-based guidance is referred to as the Canadian Updateable Model Output Statistics (UMOS) model, and its accuracy and the steps taken to remove bias from the GEM are described by Wilson and Vallée (2003). UMOS is based on data ranging from 3 to 15 h old, as it is produced after the 0000 and 1200 UTC model runs, and its data are available 3 h after these times at 0300 and 1500 UTC. UMOS guidance valid at 3-h intervals is available

for each site. Model guidance describes the following conditions: vector wind, temperature and dewpoint temperature, and precipitation occurrence and type.

To interpolate in the 0–6-h time frame, the following three steps are applied:

1) Continuous conditions (vector wind, temperature, and dewpoint temperature) are interpolated linearly from observed to model-based values.
2) Model-based guidance of precipitation is verified against the $t - 0$ h observation before its use in the 1–2-h period. If the model correctly forecast occurrence or nonoccurrence, it is used in the 1–2-h period; if not, then a persistence of the $t + 0$ h observation is used instead. In the 3–6-h period, the model-based precipitation forecast is used regardless of verification at $t - 0$ h.
3) Finally, to promote consistency, the precipitation type for each hour in the 1–5-h period is checked against the concurrent temperature. If it is obviously inconsistent, the type is changed to be consistent (e.g., if due to interpolation the temperature fell below −2°C, any rain would be changed to snow).

Note that in some cases the above steps can introduce large errors, particularly if actual conditions are about to change sharply, or if model-based forecasts are poor. Forecasters are advised to be alert to such cases

and to doubt the resulting analog forecasts. However, in most cases, these steps describe plausible conditions in the 1–6-h time frame, which in turn provide a basis for analog forecasts.

c. Fuzzy similarity-measuring algorithm

The fuzzy similarity-measuring algorithm is used to find and retrieve past cases most similar to the present case (Fig. 4). It measures the similarity of two types of variables: continuous and categorical. All of the variables routinely reported in METARs (Table 1) are continuous except for one, precipitation, which is categorical.

All of the continuous variables, except for wind speed, are compared for similarity using the fuzzy set operation shown in Fig. 3, where the dimensions of the fuzzy set are specified by the values listed in Table 2. For example, when two wind directions are compared, a difference of 10° maps to *very* similar or $\mu_{\text{direction}}(10) = 0.9$, a difference of 20° maps to *quite* similar or $\mu_{\text{direction}}(20) = 0.5$, and a difference of 40° maps to *slightly* similar or $\mu_{\text{direction}}(40) = 0.25$.

Wind speed is the only continuous variable that is compared using a second type of fuzzy operation—a fuzzy decision surface (Fig. 5). For example, if two speeds are 5 and 10 kt, they are described as *quite* similar or $\mu_{\text{speed}}(5, 10) = 0.5$, and if two speeds are 5 and 20 kt, they are described as *slightly* similar or $\mu_{\text{speed}}(5, 20) = 0.25$.

Precipitation is the only categorical variable, and it is compared using a third type of fuzzy operation—a fuzzy relation (Table 3). Thus, for example, $\mu_{\text{pcpn}}(\text{rain},$

FIG. 5. Fuzzy decision surface, μ, used to represent the degree of similarity between two values of wind speed, x_1 and x_2 in which $\mu(x_1, x_2)$ ranges from 0.0 (dark) to 1.0 (light). Generally, μ equals the quotient of the lower value divided by the higher value. When both values are near 0, the decision surface is adjusted to represent higher similarity (as shown at lower left).

rain) = 1.0 or *identical*, $\mu_{\text{pcpn}}(\text{rain, showers}) = 0.75$ or between *very* and *quite* similar, and $\mu_{\text{pcpn}}(\text{rain, drizzle}) = 0.50$ or *quite* similar. For any two types of commonly reported precipitation, a similarity measurement is achieved using a lookup table (an expanded version of Table 3).

The values in Tables 2 and 3 are based on aviation forecasters' subjectivity, obtained by asking forecasters what differences between values in weather variables they would regard as corresponding to *very*, *quite*, and *slightly* similar conditions. This is a standard approach in designing fuzzy expert systems: acquiring knowledge from experts and encoding it in fuzzy rules (Kuciauskas et al. 1998; Meyer et al. 2002). The motivation for combining fuzzy operations with the analog method is not to objectively measure similarity, nor is it to somehow

TABLE 2. Values used to calibrate fuzzy sets to measure similarity between two continuous conditions. Values correspond to operations and widths of fuzzy sets as illustrated in Fig. 3. Values are absolute for all conditions, except for the ceiling height and visibility, for which fractional values are used. Attributes not listed here are not compared (e.g., precipitation intensity, pressure tendency).

Attribute	Slightly similar	Quite similar	Very similar
Date of the year (Julian)	60 days	30 days	10 days
Hour of the day (UTC)	2 h	1 h	0.5 h
Cloud amount	4/10	2/10	1/10
Ceiling height*	¼	½	¾
Visibility*	¼	½	¾
Wind direction	40°	20°	10°
Dewpoint temperature	4°C	2°C	1°C
Dry-bulb temperature	8°C	4°C	2°C

* For example, for ceiling height, a value of ¾ implies that if the lower ceiling is 600 ft and the higher ceiling is 800 ft, then they will be described with words such as "*very* similar" and with $\mu = 0.9$.
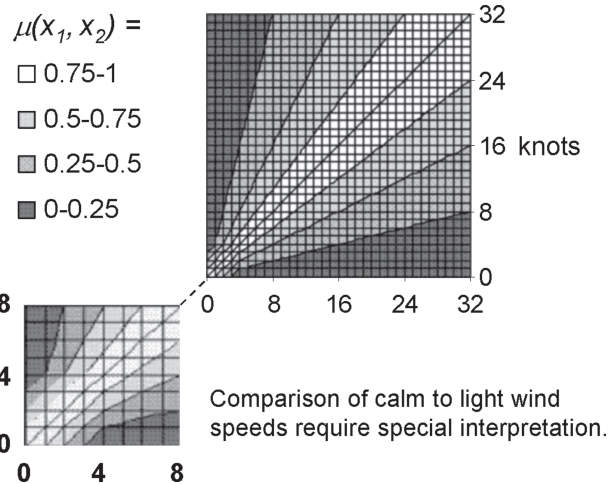
TABLE 3. Fuzzy relations used to measure similarity between precipitation types, all variables are referred to as "weather" in METARs. Only part of the lookup table is shown. The complete table has values to enable comparisons of any type of observable weather (e.g., ice pellets, snow grains, etc.). The lookup table associates similar weather types (e.g., rain resembles showers) and disassociates dissimilar weather types (e.g., drizzle does not resemble ice pellets).

	No rain	Drizzle	Showers	Rain	Snow	
No rain	1.00					
Drizzle	0.02	1.00				
Showers	0.03	0.50	1.00			
Rain	0.01	0.50	0.75	1.00		
Snow	0.01	0.05	0.05	0.05	1.00	
	No rain	Drizzle	Showers	Rain	Snow	...

optimize the analog method; rather, it is to emulate expert aviation forecasters in assessing degrees of similarity between variables that describe current weather (actual and forecast conditions) and potential past analogs, and to apply such assessments for analog forecasting.

Because the analog method itself reflects variations in independent variables (location, season, etc.), the same similarity tests and values are applied for each forecast at each site and time. The specific climatology of each site is reflected by the fact that in forecasting for the site, k-NN results are drawn from observations for only that site. Seasonal variations are reflected by a higher degree of similarity attributed to past observations that come from closer Julian dates.

Corresponding variables in the present case and approximately 350 000 past cases are compared using the fuzzy operations described above, and the most analogous cases are saved in an ordered list (Fig. 4). These analogs, the fuzzy k-NN, are used to make analog forecasts. The whole computation process takes only a small fraction of a second for each forecast. Efficiency is achieved by stopping tests and proceeding to the next candidate analog should a pair of variables measure as less similar than the least similar of the k-NN found so far. Computationally, the order of the algorithm drops quickly: $O(n^3) \rightarrow O(n)$. In other words, the time required to search an airport's archive for the k-NN is proportional to the number (n) of observations in the archive.

Referring to Fig. 4, in assessing the degree of similarity between projected times, $a(1, \ldots, 24)$, and respective potential analog times, $b(1, \ldots, 24)$, two types of similarity are assessed: the similarity of trends and conditions at respective *times zero* ($t - 1$ h and $t + 0$ h), and the similarity at respective projection times themselves ($t + 1$ h, $\ldots$, $t + 24$ h). The former has two properties: first, it permits direct comparison of the ceiling and visibility themselves, values that are not well forecast by models or MOS for $a(t + 1, \ldots, 24)$, and second, it has a diminishing influence as the projection period increases (i.e., similar conditions at respective $t - 1$ h and $t + 0$ h may be assumed to have more value for identifying analogs for $t + 1$ h than for $t + 12$ h). Therefore, for assessing the degree of similarity between current and past conditions, two methods are used: first, for the 1–6-h period, the measured similarity is determined by the minimum of both types of similarity, and second, for the 7–24-h period, the measured similarity is determined only by the similarity between conditions at the respective projection times themselves, without regard to what conditions were at the
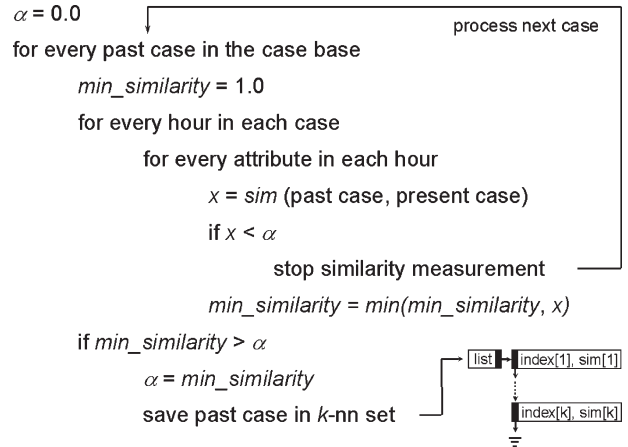

FIG. 6. Algorithm flow chart. Process identifies and orders most similar past cases according to their overall similarity with present case.

respective times zero and without regard to the actual ceiling and visibility data in $a(t - 1)$ and $a(t - 0)$.

A flow chart for the algorithm is shown in Fig. 6. The threshold for admission to the k-NN set is referred to as the α level, the lowest level of similarity among the k-NN. There are three constraints on α:

1) $0.0 < \alpha \leq 1.0$,
2) the α level is initialized to 0.0, and
3) the α level rises progressively during the climate archive search.

Thus, the computational cost of the similarity measurement decreases progressively during the climate archive search. In essence, $(1.0 - \alpha)$ is the normalized radius of a hypersphere centered on the present case that contains k-NN after climate archive traversal is complete. The search for the k-NN may be visualized as a progressively contracting hypersphere centered on the present case (Fig. 7). Points correspond to observed weather states at particular hours. Axes correspond to differences between compared variables. Differences are determined by fuzzy similarity-measuring functions, expertly tuned for a first approximation, all applied together simultaneously.

When all of the variables in two particular observations are compared, the overall similarity of the observations is taken to be the minimum of all of the similarity values, for two reasons: computationally, it is efficient and fast, and interpretatively, it ensures that all variables of the k-NN are at least as similar as the least similar variable. An example of similarity measurement between two observations using fuzzy operations is shown in Table 4. Values of the similarity between individual attributes are calculated using the operations
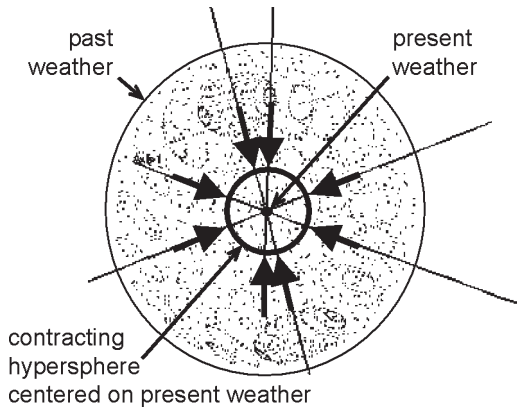
FIG. 7. A way to visualize analog forecasting using the most similar analogs.

described above and by Tables 2 and 3 and Fig. 3. The METARs would normally be written as follows:

METAR "A" 151200Z 08012KT 1SM −RA BR
  BKN008 OVC015 08/08 RMK SF8SC2
METAR "B" 251200Z 10009KT 4SM −SHRA
  BKN006 OVC015 07/07 RMK SF6SC4

d. Forecast composition

After the closest analogs (the *k*-NN) are identified, a simple method is applied for forecast composition. For the ceiling, values of the ceiling in the *k*-NN are listed in order of increasing value from lowest to highest, and the value at the 30th percentile position, x_{cig}, is applied as a deterministic prediction of the ceiling (i.e., in the *k*-NN, 30% of the observations have ceilings below or at x_{cig}, and the other 70% have ceilings higher than x_{cig}). The same method is applied to obtain a deterministic prediction of the visibility, x_{vis}. Referring to Fig. 4, this method reorders the list of *k*-NN twice: first, from

in order of decreasing similarity to in order of increasing ceiling; second, to in order of increasing visibility.

In preliminary experiments, it was initially supposed that forecasts based on the median value in the ordered list (the 50th percentile) would give accurate forecasts (Hansen 2000). However, through experimentation, it was found that using the 30th percentile gave better results, a better balance between the probability of detection (POD) of low ceilings and the false alarm ratio (FAR) of low ceilings. Because of how POD and FAR are formulated, as the percentile value used lowers, both the POD and the FAR of low ceilings rise. Referring to Table 5, the formulas are POD = $a/(a + c)$ and FAR = $b/(a + b)$.

Individual 24-h forecasts of ceiling and visibility are plotted alongside observed values for the same time period in Fig. 8. Two qualities of the forecast are apparent: forecast values often agree with observed values, particularly with respect to flight category, and forecast values sometimes vary sharply from one hour to the next. These qualities agree with the manner in which Hastie et al. (2001) characterize the *k*-NN method: "its predictions are often accurate but can be unstable."

Referring to Fig. 8, forecast conditions (GEM–UMOS guidance) were (a) light rain and snow beginning at 1200 UTC 2 December, changing to light snow at 1600 UTC and continuing through 0600 UTC 3 December; (b) light northeasterly winds shifting to light westerly winds around 1200 UTC 2 December, then strengthening gradually for the rest of the period; and (c) dewpoint temperature near 0°C through 1800 UTC lowering steadily thereafter to −9°C by 0600 UTC 3 December. Actual conditions were (a) light snow throughout the period, mixed with rain only between 0900 and 1500 UTC; (b) winds generally as forecast; and (c) dewpoint temperature generally as forecast.

Accordingly, WIND-3 forecast the ceiling and visibility to lower abruptly at 1200 UTC, consistent with the forecast onset of precipitation at that time. From 1200 UTC 2 December to 0600 UTC 3 December, there was a gradual rising trend in the ceiling and visibility, consistent with the wind shift from northeasterly to westerly and gradual drying of the air. In this forecast case,

TABLE 4. Example of calculations of the overall similarity of two METARs: $\mu_{overall}$, is the minimum of all the individual values $(\mu_1, \ldots, \mu_n)$. In this case, $\mu_{overall} = 0.25$ or *slightly* similar, due mainly to a large difference in visibility, 1 vs 4 mi.

Attribute $(a_1, \ldots, a_n)$	METAR A	METAR B	Similarity $(\mu_1, \ldots, \mu_n)$
Date of the year	15 Jul	25 Jul	0.90
Hour of the day	1200 UTC	1200 UTC	1.00
Wind direction	80°	100°	0.50
Wind speed	12 kt	9 kt	0.75
Visibility	1 SM	4 SM	0.25
Precipitation type	Rain	Rainshower	0.75
Cloud amount	8/10	6/10	0.50
Ceiling height	600	800	0.90
Dry-bulb temperature	8°C	7°C	0.95
Dewpoint temperature	8°C	7°C	0.90
$\mu_{overall}$ = minimum$(\mu_1, \ldots, \mu_n)$ =			0.25

TABLE 5. Contingency table for categorical forecasts of binary events.

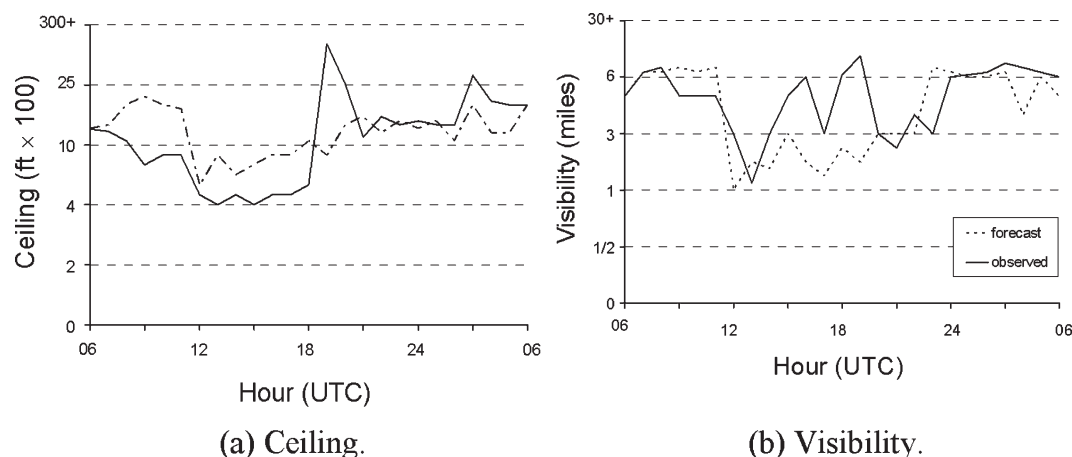|		Event observed	
		Yes	No
Event forecast	Yes	*a* (hit)	*b* (false alarm)
	No	*c* (miss)	*d* (correct rejection)

FIG. 8. Actual observations and matching 24-h analog forecasts and for Montreal International Airport (CYUL) for 2–3 Dec 2005: (a) ceiling and (b) visibility. Solid lines are actual values and dashed lines are forecast values. The x axes are time (UTC); y axes and horizontal gridlines show critical thresholds of flight categories.

because the ambient conditions were well forecast by models, the accompanying forecasts of the ceiling and visibility were quite accurately forecast by WIND-3. The following section describes the accuracy of WIND-3 based on a large number of cases.

3. Results

This section describes the accuracy of WIND-3 forecasts based on verification using the Heidke skill score (HSS) and summarizes feedback from forecasters.

a. Verification

Studies of official forecasts (TAFs) and objective forecasts have shown that persistence forecasts are a highly competitive benchmark in the 0–6-h period (Reid 1978; Dallavalle and Dagostaro 1995). Several systems have used persistence as a benchmark to measure skill in this forecast period (Wilson and Sarrazin 1989; Vislocky and Fritsch 1997; Leyton and Fritsch 2004; Jacobs and Maat 2005). In addition to reporting this type of skill, Jacobs and Maat (2005) report relative accuracy compared to both persistence and official TAFs and skill in all projection periods out to 18 h.

The accuracy of WIND-3 was verified with approximately 350 000 hindcasts for 190 stations for the period February–April 2005. HSS was calculated using a 2×2 contingency table of forecasts versus observations of two exclusive flight categories: instrument flight rules (IFR) and visual flight rules (VFR). Whereas POD and FAR each only refer to two of four values tallied in the contingency table, HSS refers to all four values and is, thus, a more comprehensive measure of forecast accuracy. With reference to Table 5, the formula for HSS is

$HSS = 2(ad - bc)/[(a + c)(c + d) + (a + b)(b + d)]$ (Wilks 2006). IFR conditions exist if ceiling is below 1000 feet or if visibility is below 3 miles; otherwise, VFR conditions exist.

WIND-3 forecasts were verified alongside two benchmark methods, persistence and current airport forecasts (TAFs). To enable fairly direct comparison, all three methods were verified using the technique described by Stanski et al. (1999). Basically, the technique uses the HSS to verify forecast conditions against actual conditions for every minute of any period of interest. Thus, for example, if a special observation made 30 min after the hour causes a TAF to switch from a "hit" to a "missed event," the TAF for that hour can be scored half as a hit and half as a missed event. Likewise, if the TAF for any period implies a sort of probabilistic forecast, say "A PROBXX B," the TAF can be scored partly based on a forecast of A and partly based on a forecast of B. Thus, for example, a TAF of "A TEMPO B" is treated as if it were 60% A and 40% B, and a TAF of "A PROB30 B" is treated as if it were 70% A and 30% B. Forecasts made by WIND-3 and forecasts made by persistence imply no sort of probability, so they are simply verified one to one against actual conditions for each minute of any period of interest.

A summary of the results for three forecast methods is shown in Fig. 9. TAF statistics are provided by the Services, Clients and Partners Directorate of Environment Canada and calculated following the method described by Stanski et al. (1999). Data consisted of in situ airport observations (current hourly METARs and historical ones from 1971 to 2004), and GEM- and UMOS-based guidance from the Canadian Meteorological Centre (CMC).

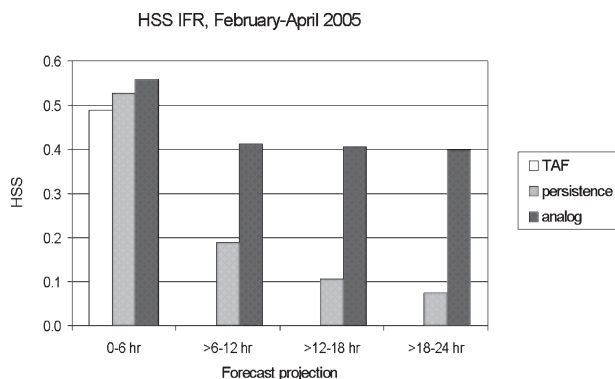HSS IFR, February-April 2005

FIG. 9. HSS for predictions of IFR conditions (ceiling <1000 ft or visibility <3 mi). Statistics are comprehensive for 190 Canadian airports for the period from Feb to Apr 2005.

WIND-3 forecasts were generally more accurate than forecasts from the competitive methods in the 0–24-h period (Fig. 9). In the 0–6 h period, the HSS of analog is higher than that of the persistence and TAFs: 0.56 compared to 0.53 and 0.49, respectively In the 7–24-h period, the HSS of the analog is markedly higher than that of the persistence: approximately 0.4 for the analog compared to 0.2 or less for the persistence (TAF statistics are not available for 7–24 h). In the second time frame, much of the skill of the analog method is due to GEM–UMOS guidance in describing ambient conditions.

b. Feedback from forecasters

Since late 2004, WIND-3 forecast guidance products have been provided in real time to operational meteorologists at MSC's two meteorological aviation centers. Forecasters have described the products as having a "high glance value," in that they allow users to quickly and easily see where conditions are most likely to change and how.

Forecasters have described WIND-3 forecast guidance as transparent and comprehensible. If a forecaster with a WIND-3 forecast wonders about its basis, he or she can examine a secondary display with more detailed information, which includes a list of all of the assumed and driving conditions (1–24-h forecasts of wind direction and speed, precipitation occurrence and type, temperature, and dewpoint), and a summary of the actual historical observations that were used to make the forecast, the individual analogs. Forecasters can assess the reliability of the forecasts by checking these conditions against other contextual forecast information (e.g., a forecast may be based on an assumption of snow and easterly winds at forecast time $t + 6$, whereas rain and westerly winds may currently appear more likely).

4. Conclusions and recommendations

A system, WIND-3, that produces deterministic forecasts of ceiling and visibility using the analog method has been described. It uses a similarity-measuring algorithm based on fuzzy logic and a 34-yr-long database of hourly airport observations. WIND-3 predictions of IFR conditions in the 0–6-h period had an HSS of 0.56, compared to an HSS of 0.53 for persistence. In the 7–24-h period, WIND-3 predictions had an HSS of about 0.40, compared to an HSS of 0.2 or less for persistence. The system has been deployed semioperationally for over a year and is regarded by forecasters as a useful addition to their set of forecasting tools.

Recommended future work

Three possible extensions of this research appear promising. First, as suggested by Leyton and Fritsch (2003) and Jacobs and Maat (2005), integrating complementary information from radar data and short-range radar-based forecasts could significantly improve the accuracy of forecasts of the ceiling and visibility. Currently, when WIND-3 inputs faulty forecasts of precipitation, the resultant forecasts of the ceiling and visibility are inevitably negatively affected.

Second, the potential for probabilistic forecasts based on analog ensembles can be explored using appropriate verification measures. Basically, such forecasts could be made by treating the distribution of ceiling values, or visibility values, as a probability density function (PDF). Thus, for example, one could expect a 10% probability of a ceiling value at or below the value marking the lowest 10% of the PDF. Such forecasts could be verified using the relative operating characteristics (ROC).

Third, the development of more sophisticated interfaces and display systems for forecasters would increase the applicability of the analog method. A more convenient and interactive interface would increase the flexibility of WIND-3. Occasionally, forecasters would like to vary the conditions that forecasts are based upon (e.g., change precipitation occurrence or timing, or alter winds), to examine the implications of alternate scenarios.

REFERENCES

Bocchieri, J. R., and H. R. Glahn, 1972: Use of model output statistics (MOS) for predicting ceiling height. *Mon. Wea. Rev.,* **100,** 869–879.

——, R. L. Crisci, H. R. Glahn, F. Lewis, and F. T. Globokar, 1974: Recent developments in automated prediction of ceiling and visibility. *J. Appl. Meteor.,* **13,** 277–288.

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. Meteor. Soc., 213–218.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1203–1211.

Glickman, T. S., Ed., 2000: *Glossary of Meteorology*. 2d ed. Amer. Meteor. Soc., 855 pp.

Hansen, B. K., 2000: Weather prediction using case-based reasoning and fuzzy set theory. M.S. thesis, Dept. of Computer Science, Technical University of Nova Scotia, 134 pp. [Available online at http://www.cs.dal.ca/~bh/thesis.]

Hastie, T., R. Tibshirani, and J. Friedman, 2001: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, 533 pp.

Jacobs, A. J. M., and N. Maat, 2005: Numerical guidance methods for decision support in aviation meteorological forecasting. *Wea. Forecasting,* **20,** 82–100.

Jacobs, W., V. Nietosvaara, S. C. Michaelides, and H. Gmoser, Eds., 2003: Inventory phase on current knowledge and requirements by users and forecasters. COST-722: Very Short Range Forecasting of Fog and Low Clouds Phase I Rep., 201 pp. [Available online at http://137.248.191.94/fileadmin/COST_media/Official_documents/Publications/publication_phase1.pdf.]

Keller, J. M., M. R. Gray, and J. A. Givens Jr., 1985: A fuzzy k–nearest neighbor algorithm. *IEEE Trans. Syst. Man Cybern.,* **15,** 258–263.

Kuciauskas, A. P., L. R. Brody, M. Hadjimichael, R. Bankert, P. M. Tag, and J. E. Peak, 1998: A fuzzy expert system to assist in the prediction of hazardous wind conditions within the Mediterranean basin. *Meteor. Appl.,* **5,** 307–320.

Leyton, S. M., and J. M. Fritsch, 2003: Short-term probabilistic forecasts of ceiling and visibility utilizing high-density surface weather observations. *Wea. Forecasting,* **18,** 891–902.

——, and ——, 2004: The impact of high-frequency surface weather observations on short-term probabilistic forecasts of ceiling and visibility. *J. Appl. Meteor.,* **43,** 145–156.

Lund, I. A., and D. D. Grantham, 1979: Estimating recurrence probabilities of weather events. *J. Appl. Meteor.,* **18,** 921–930.

——, and P. Tsipouras, 1982: A model for estimating persistence probabilities. *J. Appl. Meteor.,* **21,** 988–994.

Martin, D. E., 1972: Climatic presentations for short-range forecasting based on event occurrence and reoccurrence profiles. *J. Appl. Meteor.,* **11,** 1212–1223.

MetEd, cited 2006: Operational models matrix: Characteristics of operational NWP models. [Available online at http://meted.ucar.edu/nwp/pcu2.]

Meteorological Service of Canada, cited 2006: MANAIR–Manual of Standards and Procedures for Aviation Weather Forecasts. Meteorological Service of Canada, Environment Canada. [Available online at http://www.msc-smc.ec.gc.ca/msb/manuals/manair.]

Meyer, M. A., K. B. Butterfield, W. S. Murray, R. E. Smith, and J. M. Booker, 2002: Guidelines for eliciting expert judgment as probabilities or fuzzy logic. *Fuzzy Logic and Probability Applications: A Practical Guide,* T. J. Ross, J. M. Booker, and W. J. Parkinson, Eds., Society for Industrial and Applied Mathematics, 105–121.

National Climate Data and Information Archive, cited 2006: Documentation for the Digital Archive of Canadian Climatological Data (surface) identified by element. National Climate Data and Information Archive, Meteorological Service of Canada, Environments Canada. [Available online at http://climate.weatheroffice.ec.gc.ca/prods_servs/documentation_index_e.html.]

Reid, J. D., 1978: Verification of ceiling and visibility forecasts using the ranked probability score. *Atmos.–Ocean,* **16,** 177–186.

Riordan, D., and B. K. Hansen, 2002: A fuzzy case–based system for weather prediction. *Eng. Intelligent Syst.,* **3,** 139–145.

Stanski, H., A. Leganchuk, A. Hanssen, D. Wintjes, O. Abramowski, and J. Shaykewich, 1999: NAV CANADA's TAF amendment response time verification. Preprints, *Eighth Conf. on Aviation, Range, and Aerospace Meteorology,* Dallas, TX, Amer. Meteor. Soc., 63–67.

Stoelinga, M. T., and T. T. Warner, 1999: Nonhydrostatic, meso-beta-scale model simulations of cloud ceiling and visibility for an East Coast winter precipitation event. *J. Appl. Meteor.,* **38,** 385–404.

Stutchbury, J. F., and R. L. Hawkes, 1974: Wind velocity correlations with low ceilings and visibilities in terms of defined potential occurrences at Goose Airport. Dept. of the Environment TEC 808, Atmospheric Environment Service, Downsview, ON, Canada, 31 pp.

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.

Whiffen, B., 1993: FTGEN - An automated FT production system. Preprints, *Fifth Int. Conf. on Aviation Weather Systems,* Vienna, VA, Amer. Meteor. Soc., 327–330.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2d ed. Academic Press, 648 pp.

Wilson, L. J., and R. Sarrazin, 1989: A classical-REEP short-range forecast procedure. *Wea. Forecasting,* **4,** 502–516.

——, and M. Vallée, 2003: The Canadian Updateable Model Output Statistics (UMOS) system: Validation against perfect prog. *Wea. Forecasting,* **18,** 288–302.