Intercomparison of In-Flight Icing Algorithms. Part I: WISP94 Real-Time Icing Prediction and Evaluation Program

GREGORY THOMPSON, ROELOF T. BRUINTJES, BARBARA G. BROWN, AND FRANK HAGE

Research Applications Program, National Center for Atmospheric Research, Boulder, Colorado

ABSTRACT

The purpose of the Federal Aviation Administration's Icing Forecasting Improvement Program is to conduct research on icing conditions both in flight and on the ground. This paper describes a portion of the in-flight aircraft icing prediction effort through a comprehensive icing prediction and evaluation project conducted by the Research Applications Program at the National Center for Atmospheric Research. During this project, in-flight icing potential was forecast using algorithms developed by RAP, the National Weather Service's National Aviation Weather Advisory Unit, and the Air Force Global Weather Center in conjunction with numerical model data from the Eta, MAPS, and MM5 models. Furthermore, explicit predictions of cloud liquid water were available from the Eta and MM5 models and were also used to forecast icing potential.

To compare subjectively the different algorithms, predicted icing regions and observed pilot reports were viewed simultaneously on an interactive, real-time display. To measure objectively the skill of icing predictions, a rigorous statistical evaluation was performed in order to compare the different algorithms (details and results are provided in Part II). Both the subjective and objective comparisons are presented here for a particular case study, whereas results from the entire project are found in Part II. By statistically analyzing 2 months worth of data, it appears that further advances in temperature and relative-humidity-based algorithms are unlikely. Explicit cloud liquid water predictions, however, show promising results although still relatively new in operational numerical models.

1. Introduction

The need for better aircraft icing forecasts was reemphasized by the crash of an ATR72 commuter airplane near Roselawn, Indiana, on 31 October 1994 (NTSB 1996). The ongoing Winter Icing and Storms Project (WISP) has a goal of improving forecasts of icing conditions for the aviation community (Rasmussen et al. 1992). One portion of this effort is devoted to the development of automated icing algorithms to be used in conjunction with present numerical models. Recently, a few diagnostic icing algorithms of this type have been developed into a national-scale product (Schultz and Politovich 1992; Forbes et al. 1993). With these algorithms, temperature and relative humidity values provided by nationally available gridded numerical model data are used to determine expected regions of icing. The numerical model output is compared to preset temperature and humidity thresholds in an attempt to infer cloudy environments in appropriate temperature ranges for icing (0° to −20°C). Since operational models have only recently begun to predict explicitly the existence of clouds (i.e., cloud liquid water and cloud ice), the thresholds are used to diagnose cloudy environments and therefore icing potential. Although a few studies have been conducted to evaluate some algorithms, there has not been a coordinated effort to intercompare these algorithms.

As part of the WISP94 field program, a comprehensive evaluation program was conducted to evaluate a set of diagnostic icing algorithms applied to different numerical models. The program, WISP Real-time Icing Prediction and Evaluation Program (WRIPEP), ran concurrently with the WISP 1994 field project from 25 January to 25 March. The overall purposes of WRIPEP were to evaluate (in real time and postanalysis) the present icing algorithms as they are applied to the most up-to-date numerical forecast models, and to conduct a near-real-time verification exercise using pilot reports (PIREPs) of icing and measurements from the research aircraft during the 1994 WISP field program. Specific objectives of WRIPEP were

1) to allow real-time subjective evaluation of icing algorithms by scientists and forecasters at the National Center for Atmospheric Research (NCAR) and the National Aviation Weather Advisory Unit (NA-WAU);
2) to perform a comprehensive, objective verification and comparison of the different algorithms;

Corresponding author address: Gregory Thompson, National Center for Atmospheric Research, Research Applications Program, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: gthompsn@ucar.edu

3) to compare and evaluate explicit liquid water content model calculations with icing derived from the temperature- and humidity-based algorithms and with observations; and

4) to study the effects of different horizontal and vertical grid resolutions on the output icing diagnostic.

To meet these objectives, WRIPEP consisted of a real-time display for visualizing the icing forecasts and a statistical evaluation used to evaluate the forecast quality of a multitude of model–algorithm couplings.

This paper describes WRIPEP and sample results of the project. The numerical models used in this study are presented in section 2. Section 3 provides an overview of the icing algorithms, while the WRIPEP display is described in section 4. A case study day is presented in section 5 along with a small sampling of statistical results. A complete discussion of the statistical procedures and an in-depth evaluation of the results for the entire WRIPEP period are provided in Part II, a companion paper by Brown et al. (1997), hereafter referred to as B97. Finally, a few remarks and general conclusions are provided in section 6.

2. Numerical models

The primary models used to calculate icing potential were the Eta and MAPS modeling systems run at the National Center for Environmental Prediction (NCEP, formerly the National Meteorological Center). The Eta Model is a recent NCEP development; one of its fundamental aspects is the incorporation of step topography and the vertical eta coordinate (Mesinger et al. 1988). The Mesoscale Analysis and Prediction System (MAPS) (Benjamin et al. 1991) was developed at the National Oceanic and Atmospheric Administration's (NOAA) Forecast Systems Laboratory (FSL) and now runs operationally at NCEP as the Rapid Update Cycle (RUC) (Benjamin et al. 1994). In addition, the fifth-generation Pennsylvania State University–NCAR mesoscale model (MM5) (Dudhia 1993) was run at NCAR in real time during the field program and has been utilized in this study as well; however, these results will not be presented here.

As part of WRIPEP, two configurations of the Eta Model were utilized. The first was the 80-km horizontally spaced version with vertical information at 50-mb increments (interpolated from the model's native vertical coordinate) for a total of 19 levels. The second was a 40-km, 38-level, experimental version with data on the native horizontal and vertical coordinate systems. WRIPEP thus provided an excellent opportunity to investigate the effects of model resolution on different icing algorithms. The 80-km version was available twice daily with forecasts to 48 h, while the 40-km version was provided only at 0000 UTC daily with forecasts to 36 h. One additional advantage of the 40-km version

was the inclusion of cloud liquid water and cloud ice parameterizations as described by Zhao et al. (1997).

The MAPS model data were provided by FSL for postanalysis studies; however, NCEP's version, the RUC, was run in an experimental mode and was occasionally available during the field program. This model has 60-km nominal horizontal spacing with a hybrid vertical coordinate (lowest levels are terrain following, then transform to isentropic surfaces near the middle of the troposphere) that provided 25 vertical levels. MAPS model data were provided at 3-h intervals and contained forecasts up to 6 h in length except at 0000 and 1200 UTC when forecasts to 12 h were available.

3. Algorithms

Currently, the National Weather Service's (NWS) Aviation Weather Center (AWC, formerly NAWAU) in Kansas City provides icing forecasts for the contiguous United States in the form of airmen's meteorological information (AIRMETs) and significant meteorological information (SIGMETs). AIRMETs are issued every 6 h by AWC at 0145, 0745, 1345, and 1945 UTC and contain forecasts of icing of moderate or greater intensity for the following 6 h unless amended by another, unscheduled, AIRMET. AIRMETs are created using many weather data sources including satellite imagery, model-predicted temperature and humidity fields, and existing pilot reports of icing. SIGMETs are usually issued for known severe icing as necessary based on current pilot reports. Both are polygon in shape (described by a series of navigation aids across the United States) in the horizontal and encompass a range of altitude in the vertical, usually from the freezing level to a specified altitude. Direct comparison between AIRMETs and the automated algorithms is difficult, because an AIRMET must span a 6-h time period while the model–algorithm is valid at a single time.

Besides the ''official'' NAWAU-forecasted AIRMETs, WRIPEP included three diagnostic, automated icing algorithms described in the sections below. The algorithms were chosen for use in WRIPEP because they are currently used operationally. These algorithms are in various stages of development and testing by several organizations cooperating in WRIPEP. Because model-predicted relative humidities (RH) are often less than 100% in cloudy environments, all of the automated algorithms tested in WRIPEP use humidity thresholds less than 100%. However, in the real atmosphere, icing cannot exist in subsaturated environments (RH < 100%), except within precipitation or a decaying cloud. The reason for this discrepancy is that subgrid-scale features that are not resolved by the models and errors in the numerical model predictions require a lower than 100% RH threshold to detect/predict clouds. These algorithms are best used as first-guess icing algorithms because of their simplicity. They do not contain sophisticated microphysical relationships. Instead, they utilize model-

TABLE 1. RAP algorithm temperature (T) and relative humidity (RH) thresholds and corresponding icing categories.

Icing category	Temperatue range	Relative humidity threshold
Freezing rain	$T \leq 0°C$	RH $\geq$ 80% (with RH $\geq$ 80% above $T > 0°C$)
Stratiform	$-12 \leq T \leq 0°C$	RH $\geq$ 85% (with RH $\leq$ 85% above $T < -12°C$)
Unstable	$-20 \leq T \leq 0°C$	RH $\geq$ 56% (with max RH $\geq$ 65% within unstable layer)
General	$-16 \leq T \leq 0°C$	RH $\geq$ 63%

predicted RH to diagnose clouds and refine their search to temperatures that are appropriate for icing conditions. The use of RH as a determiner of clouds is quite crude and some clear-sky regions will be diagnosed as clouds (and icing when within the proper temperature range). Also, some diagnosed clouds could in fact be ice or snow filled since the algorithms do not discriminate between liquid and ice clouds. Nonetheless the algorithms are presented here because they are used operationally and can be useful to those without access to microphysical model data.

In contrast to the T–RH algorithms, the 40-km version of the Eta Model and all resolutions of the MM5 model also contain water and ice parameterizations that prognose cloud liquid water (CLW) and ice fields. CLW, too, was used to predict locations of anticipated aircraft icing and is verified statistically in B97.

a. RAP icing scheme

The RAP algorithm, developed by the NCAR Research Applications Program is a continually evolving algorithm refined from earlier versions by Schultz and Politovich (1992) and Forbes et al. (1993, hereafter referred to as Forbes). The algorithm currently consists of four categories of icing, shown in Table 1, including three adapted from Forbes (General, Unstable, and Freezing Rain), plus a new category labeled Stratiform. The algorithm's thresholds were adjusted several times throughout and after the WRIPEP period in an attempt to improve statistical verification results. Unfortunately, this goal was unattainable and the current thresholds (Table 1) were selected by using the means and standard deviations of the observational data provided in Forbes. These observations were based on sounding data in the vicinity of PIREPs. For example, using sounding data, Forbes found a mean temperature of $-8.4°C$ and standard deviation of $7.2°C$ at the level of reported aircraft icing; from this, the threshold of $-16°C$ was formed. Furthermore, they found a mean relative humidity of 82% with standard deviation 19%; hence, the threshold of 63% was formed.

The four categories of the RAP algorithm were chosen in order to provide forecasters with readily distinguishable physical reasons for icing conditions. If the forecasters know that the physical reasoning behind the icing forecasts is incorrect, then they can more readily adjust the icing diagnostic accordingly. For example, if the algorithm predicts unstable icing in a region that the forecaster clearly knows is not unstable (which can often

occur when model-predicted frontal position is ahead of or behind the actual position), then the forecaster can choose to ignore or adjust the predicted region. This provides a more useful forecasting tool than a single-category yes/no prediction with no clues as to its cause. It is strongly emphasized that the categorical output of this algorithm provides no indication of the icing type or severity. Furthermore, all categories provide an "icing potential" forecast with no probability assigned or implied. At the moment, the authors feel icing type, severity, or probability forecasts are not possible with the current algorithm design.

The general portion of the algorithm is nearly the same as the Schultz and Politovich (1992) algorithm. This category attempts to infer large-scale cloud regions with a simple temperature (T) check in a set interval ($-16°$ to $0°C$) and a test for RH greater than some threshold (63% in this case). Forbes suggested $-20° \leq T \leq +2°C$ and RH $\geq$ 56%, which were used initially in WRIPEP but were found to overforecast regions of icing to a large extent (discussed further in section 6).

The unstable portion of the algorithm is similar to the general category in that it attempts to infer large-scale cloud regions; however, it does so with the restriction that the region be conditionally unstable or conducive to convective cloud development. Based on evaluations of rawinsonde data in the vicinity of PIREPs of icing, Forbes concluded that as many as 80% of the PIREPs in dry environments (RH $\leq$ 80%) were in conditionally unstable environments (lapse rate less than moist adiabatic). For this reason the unstable category has a lower relative humidity threshold (RH $\geq$ 56%) than the general category but requires a higher RH value (RH $\geq$ 65%) within the conditionally unstable layer below the level in question. In this manner, subsaturated environments that may contain icing due to convective turrets rising from lower levels are detected. Because of the low RH threshold, this scheme overforecasts icing regions but is intended to indicate regions of intermittent icing. This may be the case when a plane is flying through convective elements of a lower-level cloud deck, such as stratocumulus or, perhaps, convective elements embedded within a stratiform cloud.

The stratiform portion of the algorithm is based on observations by Huffman and Norman (1988), Pobanz et al. (1994), and Vilcans (1989) and attempts to identify regions of "warm" stratus clouds with temperatures between $0°$ and $-12°C$. Relative humidty values greater than 85% are required within the $0°$ to $-12°C$ range and RH values less than 85% are required at tempera-
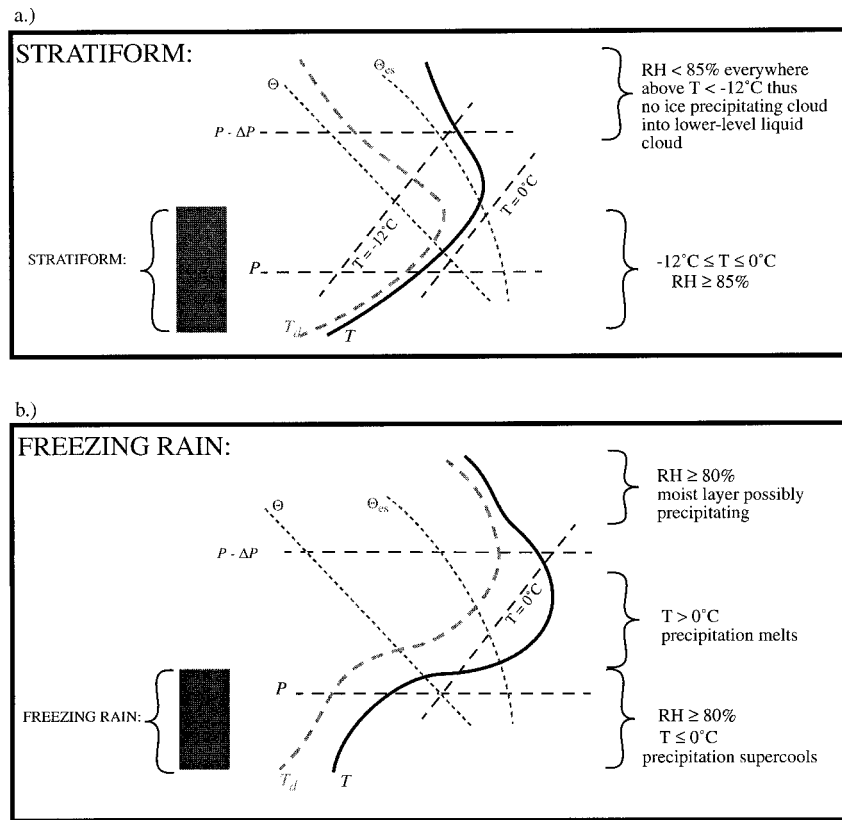
FIG. 1. Graphical schematic (portion of skew T–log$p$ diagram) depicting the (a) stratiform and (b) freezing rain portions of the RAP icing algorithm. The solid black line represents a temperature profile with height while the dashed, thick gray line represents the dewpoint temperature profile with height. Temperature and relative humidity thresholds from Table 1 are also shown on the right of the diagrams.

tures less (usually vertically higher) than $-12°C$. Huffman and Norman's study found a high frequency of this occurrence—water saturated, relatively warm low levels capped by very dry air aloft—using rawinsonde data and attributed it to the inefficiency of ice nuclei at temperatures greater than $-10°C$. Also, Vilcans found 17 events over a 23-yr period in which these conditions were present when freezing drizzle was being reported at the surface in Denver, Colorado. In a different study, Pobanz et al. found relatively warm cloud tops (temperatures $> -15°C$), thermodynamic stability, and dynamic instability (Richardson number < 1) correlated well with freezing drizzle events. The stratiform scheme criteria are intended to screen out regions where upper-level clouds may exist, which could precipitate ice into the low-level cloud and scavenge out the liquid (Politovich and Bernstein 1995). The graphical depiction of this scenario in Fig. 1a shows a typical profile associated with shallow postfrontal clouds associated with Arctic fronts on the east side of the Rocky Mountains.

Finally, the freezing rain portion of the algorithm attempts to mimic freezing rain events typical in the midwestern and northeastern United States during winter months. In this case, a layer of air with temperatures greater than $0°C$ overlies a layer of air with temperatures less than $0°C$ (Fig. 1b), a typical case of "isentropic lift" common to the north of warm or stationary fronts. The RH above the $T > 0°C$ layer ("warm nose") must be $\geq 85\%$, indicating an assumed precipitating cloud (either by coalescence or ice mechanism); while the relative humidity in the $T < 0°C$ layer must be $\geq 80\%$ indicating that enough moisture should exist such that the falling precipitation will not evaporate. Two deficiencies currently exist with this technique: 1) the total area of the warm nose is not used to check that frozen precipitation falling from higher levels has sufficient time to melt, and 2) the relative humidity between the assumed precipitating cloud and the warm nose is not checked to see if it has sufficient time to evaporate—in reality, it could evaporate before reaching the melting level. To make this algorithm more complete, future installments of this portion of the algorithm should include checks for these two items.

b. NAWAU icing scheme

The NAWAU algorithm, developed at NAWAU in Kansas City, is also a refinement of the algorithm de-

TABLE 2. NAWAU algorithm temperature (T), relative humidity (RH), and boundary layer thresholds and corresponding icing categories.

Icing category	Temperature range	Relative humidity threshold	Layer condition
Category 2*	$-14 \leq T \leq -1°C$	RH $\geq 75\%$	above boundary layer (900 m AGL)
Category 2*	$-20 \leq T \leq 0°C$	RH $\geq 86\%$	within boundary layer
Category 1*	$-19 \leq T \leq 0°C$	RH $\geq 60\%$	

* Lowered by one category if downslope winds less than -5 cm s^{-1} are present.

veloped by Schultz and Politovich (1992) and is used operationally as guidance (R. Olson 1993, personal communication) for AIRMET preparation. This algorithm includes two categories that attempt to classify the possibility of icing. The first category is considered to have a smaller probability of icing than the second category. The thresholds for this scheme, listed in Table 2, were determined by real-time forecaster comparisons with pilot reports at NAWAU. Besides the thresholds, this scheme differs from the Schultz and Politovich algorithm by reducing areas of icing when orographic downslope flow exists. It does this by decreasing the icing category by one if downslope flow greater than 5 cm s^{-1} exists within 500 m of the surface.[1] An initial test of this feature of the algorithm shows that its contribution is negligible to the statistical measures used in WRIPEP; however, more testing may be required since the 500-m requirement only encompasses one or two model levels and thus may be too restrictive.

c. AFGWC RAOB icing scheme

An algorithm developed by the Air Force Global Weather Central (Knapp 1992) for guidance of their flight operations was also included in WRIPEP. This icing algorithm was originally written for use with rawinsonde data but was tested in WRIPEP on model data. The algorithm, shown in Table 3, has three temperature categories with dewpoint depression checks and stability dependencies devised to distinguish individually all types and severities of icing. This scheme was previously evaluated statistically versus other icing algorithms; however, its individual classifications were

found to have little or no forecast skill (Cornell et al. 1995; C. Bjorkman 1995, personal communication). For the WRIPEP exercise, all the different serverities/types were collapsed into a single yes/no forecast for statistical analysis, although the visual display utilized all categories.

4. Display

To evaluate subjectively multiple algorithms and models simultaneously, software was written to display and visualize the data in real time or postanalysis. A limited example of the display capabilities is shown in Fig. 2, which shows the RAP icing algorithm output produced using 12-h RUC model forecast variables (temperature and relative humidity). The valid time of the icing prediction is 0000 UTC 28 January 1994 and positive and negative pilot reports of icing are displayed directly on top of the predicted icing. Caution must be exercised when viewing this figure since the PIREPs displayed are for all levels in the atmosphere while the icing is for a constant pressure level of 925 hPa.

One powerful aspect of the real-time display not conveyed in Fig. 2 is the ability to display three model or algorithm results simultaneously. Figure 2 simply shows one portion of the actual display. The real-time display consisted of four X windows and a graphical user interface (GUI) for interfacing with the user's requests. Three of the windows were used to show model-generated forecasts while the fourth showed symbolic products such as PIREPs or AIRMETs and SIGMETs. This window was intended for verification or observational data and could also be used to display surface aviation observations or other icon or polyline products. The three model display windows could contain different models and/or different algorithms or model variables as determined by user selection through the GUI. The capability also existed to create vertical cross sections along selectable flight paths to determine aircraft icing

---

[1] The downslope component of the flow is computed by summing the u wind component times the change in terrain elevation over the x direction, and the v wind component times the change in terrain elevation over the y direction.

TABLE 3. Air force algrithm temperature (T, °C), dewpoint depression (ddp, °C), and lapse rate (Γ, °C/1000 ft) thresholds and corresponding icing intensities and types.

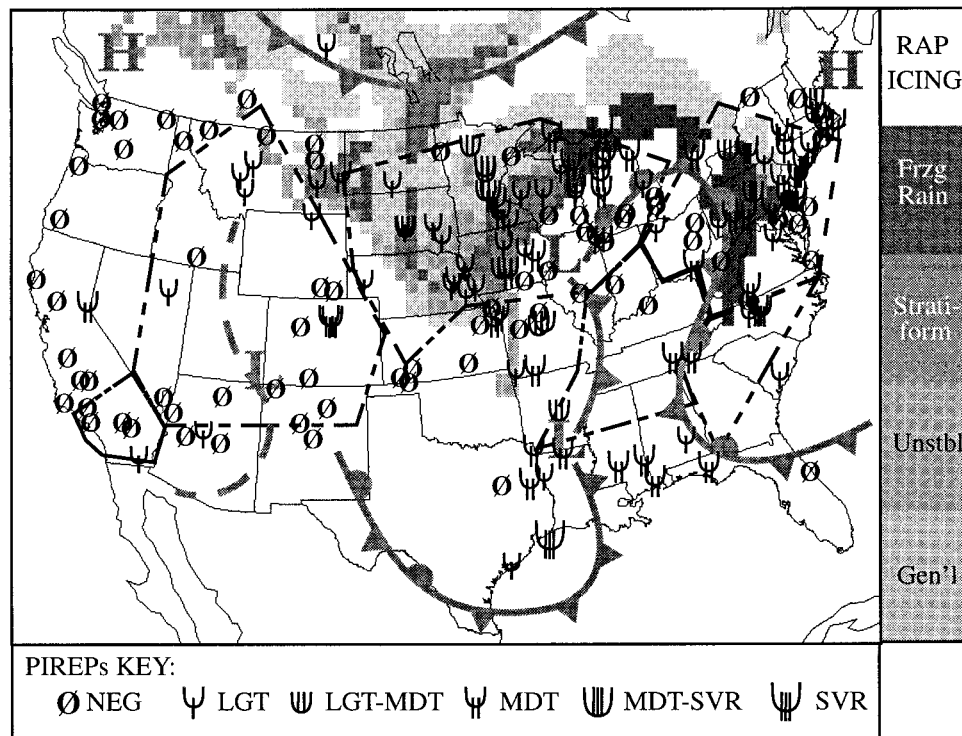| Temp	$-8 < T \leq 0$				$-16 < T \leq -8$				$-22 < T \leq -16$
Dewpoint depression	ddp ≤ 1		$1 < $ ddp ≤ 2		ddp ≤ 1		$1 < $ ddp ≤ 3		ddp ≤ 4
Lapse rate	$\Gamma \leq 2$	$\Gamma > 2$	$\Gamma \leq 2$	$\Gamma > 2$	$\Gamma \leq 2$	$\Gamma > 2$	$\Gamma \leq 2$	$\Gamma > 2$	N/A
Icing	LGT RIM	MDT CLR	TRC RIM	LGT CLR	MDT RIM	MDT MXD	LGT RIM	LGT MXD	LGT RIM

FIG. 2. Example of a WRIPEP display window showing the predicted icing at 925 hPa valid at 0000 UTC 28 January 1994. The icing forecast was created using the RAP algorithm in conjunction with the 12-h forecast RUC model variables. The different categories of the RAP algorithm are shown in gray shades along with pilot reports of icing (both positive and negative) at all levels and between 2100 and 0300 UTC. AIRMETs issued at 1945 UTC are also shown as the thick dashed lines.

hazards en route. The symbolic products (PIREPs) could be overlaid on any other window to perform a real-time, subjective evaluation of the icing predicted by the model–algorithm couplet. Switching between the different algorithms could be accomplished by a simple click of a computer mouse button. In this manner, one could easily view the output of various model–algorithm combinations and subjectively evaluate their performance. The WRIPEP software was written to enable additional algorithms to be incorporated easily. For example, several diagnostic turbulence algorithms were included in the WRIPEP display capability; other variables, such as ceiling and visibility, are planned to be included in future tests.

Because of the amount of numerical model data, the WRIPEP software requires reasonably high amounts of computer memory (at least 32 MB of RAM) and even more hard disk space (500 MB is typical). The machines utilized at both NCAR and NAWAU were UNIX workstations. The model datasets were transferred from NCEP to NCAR via the Internet and, once on site, were transformed to a format that the display expected. The WRIPEP software contact their data sources via the Internet using client–server protocols. This capability allows users to view datasets that may be impractical to copy to local computers.

5. Results

The purpose of this section is to show how the WRIPEP objectives listed in the introduction were satisfied. Examples of the display using a case study help illustrate how forecasters could obtain a subjective evaluation of the icing algorithms' performance. This satisfies objective (i) by subjectively comparing the algorithm forecasts against the only available observations, PIREPs. An example of a 24-h forecast of cloud liquid water from the Eta Model is also shown and, when used in conjunction with the above, satisfies objective (iii). Following this section is a discussion of the statistical analysis performed for the case study that exemplifies objective (ii). This full objective is met by analyzing the entire seasons' data (which is done in B97) as opposed to the single case study presented here. The final objective concerning the effects of horizontal and vertical resolution will be addressed further in another study.

a. Case study: 28 January 1994

At 0000 UTC 28 January 1994, a complex surface weather pattern in place over the continental United States provided an interesting icing case. An Arctic front (see Fig. 2) was moving south across the Canadian prai-
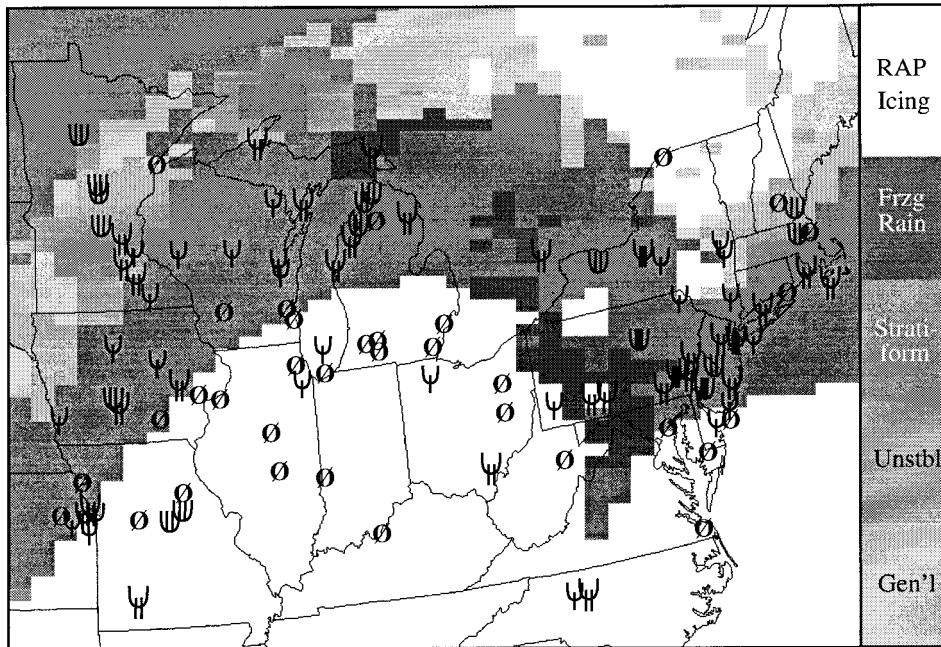
FIG. 3. The 24-h icing forecast using the RAP icing algorithm and data from the 40-km experimental mesoscale Eta Model valid at 0000 UTC 28 January 1994. PIREPs are from the ground to 1.2 km.

rie and was approaching the U.S. border. Meanwhile, a strong high pressure center located near Maine was causing cold-air damming east of the Appalachians from New York to Alabama. Also, two weak surface lows located over the Mississippi Valley were moving slowly eastward along with a trailing cold front extending south to the Gulf of Mexico, which arched westward as a stationary front in southwest Texas. Figure 2 shows an icing forecast using the RAP algorithm and the official icing forecast, the NAWAU-issued AIRMETs from 1945 UTC 27 January (valid until 0200 UTC 28 January).

Precipitation reports at this time indicated a very broad region of light rain from Michigan to Mississippi eastward to the Carolina coast. However, to the north, there were widespread reports of freezing precipitation within the cold dome of air pushed up against the Appalachian Mountains. There was also widespread freezing precipitation north of the surface low (in Illinois) along the southern Great Lakes. Farther north and west, the freezing precipitation changed to all snow from Minnesota to South Dakota and Nebraska. Among the areas reporting freezing precipitation, freezing drizzle was the dominant type except for a few embedded reports of ice pellets and freezing rain.

Figure 3 shows a real-time 24-h forecast of predicted icing using the RAP algorithm and numerical model output from the 40-km experimental Eta Model. The plot is valid at 0000 UTC 28 January 1994 at approximately 600 m above mean sea level (MSL). The stratiform portion of the algorithm dominates the forecast region, which is not surprising considering the synoptic

situation of cold-air damming along the Appalachians and also the shallow cloud north of the surface low. Some embedded regions of freezing rain are predicted in western Maryland and Pennsylvania. By viewing the actual pilot reports of icing between 2100 UTC 27 January and 0300 UTC 28 January, one can see that the RAP algorithm encompassed most of the reported icing. The icing prediction shown in Fig. 3 is for the eighth Eta Model level (about 600 m MSL) while the PIREPs shown are from the ground up to 1.2 km MSL.

The 24-h forecast icing using the NAWAU and air force algorithms, shown in Fig. 4, produced results similar to the RAP algorithm. In fact, all three algorithms nearly match on a gridpoint by gridpoint basis. A breakdown of the icing types/severities is also shown in the figure as predicted by the air force algorithm and subjectively appears to have little skill. Because of the air force's more restrictive thresholds (especially in terms of humidity), its prediction is entirely contained within that of both RAP's and NAWAU's. Though not very noticeable in this case study, much smaller areas predicted by the air force algorithm compared to RAP and NAWAU were typical throughout much of the WRIPEP project.

As stated earlier, the experimental version of the Eta Model contained a CLW parameterization. A 24-h forecast of this model variable is shown in Fig. 5. CLW amounts as high as 0.5 g kg^{-1} were predicted in the mid-Atlantic and western Great Lakes regions. Again, comparing the locations of actual icing reports, one can see the model-predicted cloud liquid water field encompassed nearly all of the reports. Although actual cloud
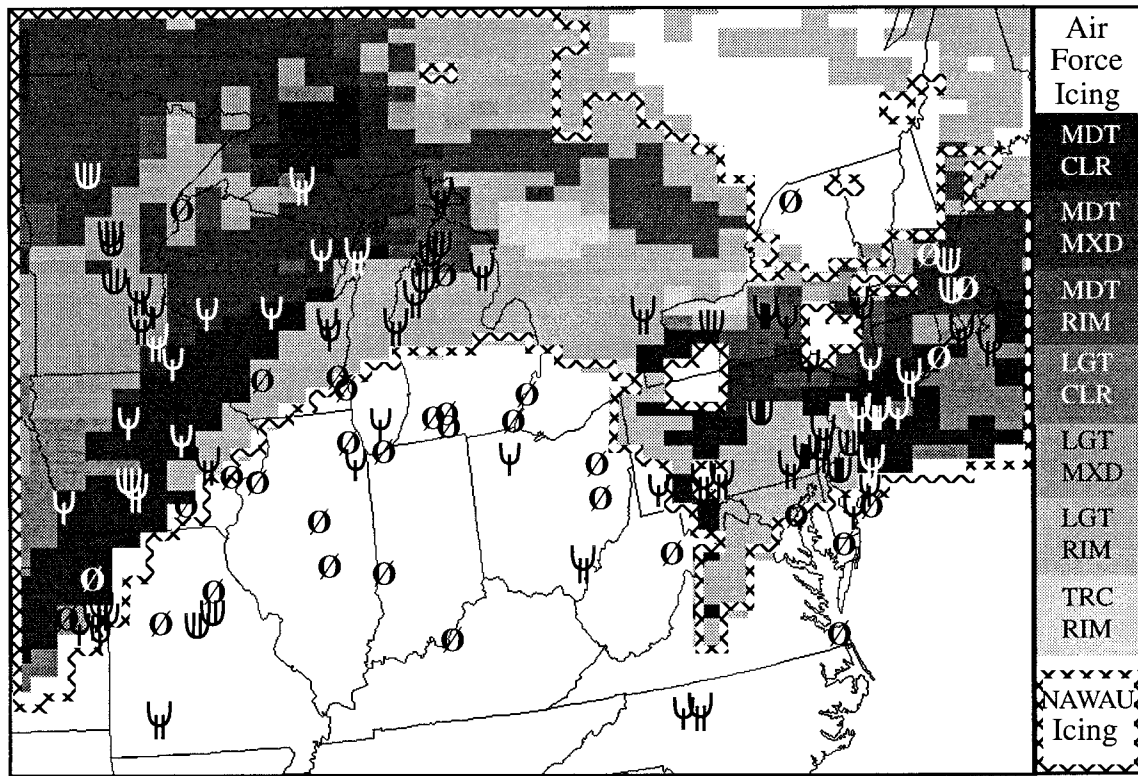
FIG. 4. Same as Fig 3 except grayscale denotes icing forecast using the air force algorithm and the thick hatched outline denotes icing forecast by the NAWAU algorithm.
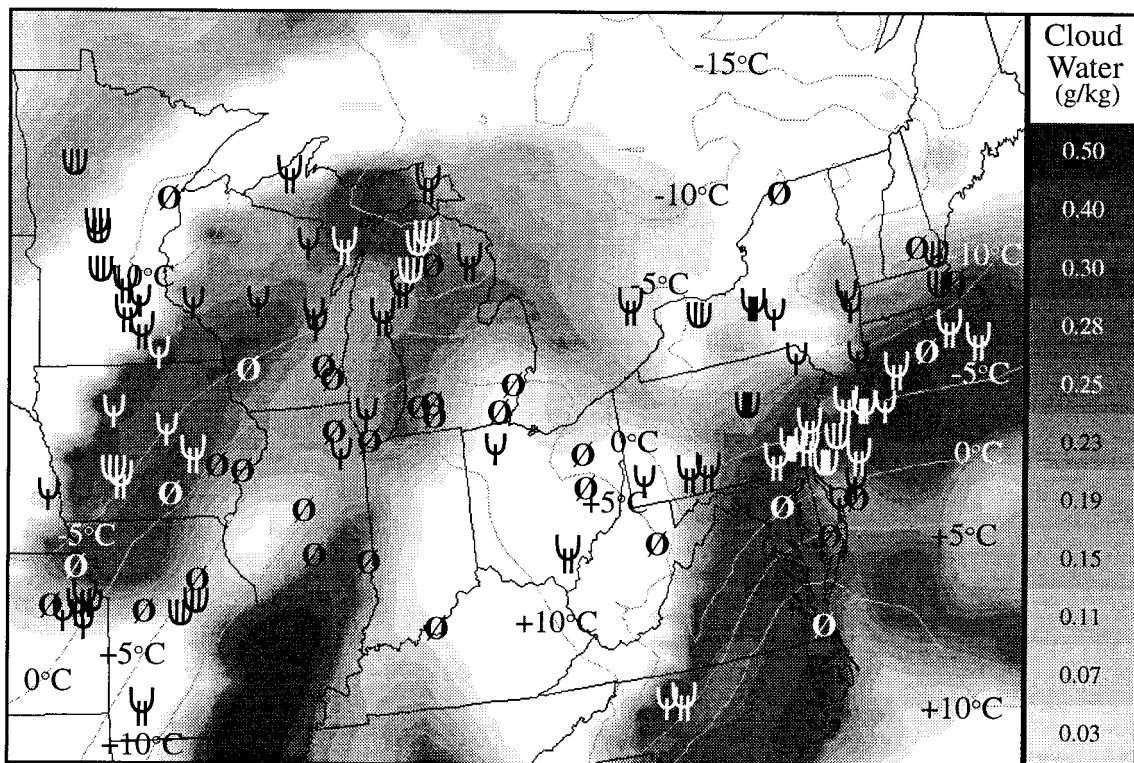

FIG. 5. Same as Fig 3 except gray shading depicts cloud liquid water (g kg^{-1}) and contours represent temperature (°C).

liquid water amounts cannot be directly verified, subjectively this appears to be a reasonably accurate forecast, especially considering it is 24 h in advance.

b. Statistics

WRIPEP results included an immense quantity of statistical verification data. A sample of the types of results obtained is presented here for the 28 January case study day. A complete discussion of the statistical analysis and results for the entire WRIPEP period are presented in B97.

As in Forbes, verification of the icing algorithms was performed using PIREPs of icing. The PIREPs used for the WRIPEP verification were obtained from the NWS family of service's Domestic Data Service communication circuit. They originate from aircraft as voice-transmitted PIREPs and are received by Flight Service Stations and Air Route Traffic Control Centers where they are entered into an NWS communication gateway. Mandatory elements of PIREPs include location information, time, flight level, and aircraft type. Other, optional elements are sky conditions, visibility and present weather, temperature, winds, turbulence, icing, and remarks. Very few PIREPs contain all of the above elements. For the WRIPEP verification, PIREPs containing either positive or negative icing reports were used. PIREPs have numerous inherent errors and biases (Schwartz 1996), but they remain the only source of observational data available routinely on the national scale.

PIREPs were compared to the diagnostic model output in order to determine the probability of detection (POD) (Wilks 1995). POD can be interpreted as the proportion of observed icing PIREPs that were correctly forecast (Brown et al. 1997). A POD equal to 1.0 means that every observed PIREP of icing was forecast by the algorithm to have icing conditions, while a POD < 1.0 indicates that some PIREPs were outside of algorithm/forecast regions. For the case study presented here, two sets of POD statistics were computed. For the first statistic, "POD_{All}," positive reports of all severities contributed to the computation of POD. For the second statistic, "POD_{MOG}," the trace and light icing reports were neglected while light-to-moderate through Severe were included. POD_{MOG} is computed in order to ignore the less-than-moderate icing reports, which are not considered to be an aviation hazard. Taking this approach assumes that pilots' reports of severity can be broadly categorized. Although a number of subjective factors influence pilots' severity reports (pilot's experience, aircraft response, and visual cues to name a few), the broad moderate-or-greater (MOG) category should be representative of more severe icing conditions for most pilots and aircraft. POD_{All} provides an indication of how the algorithms perform in an overall sense, while POD_{MOG} evaluates how well they capture the more severe icing reports.

TABLE 4. Statistics for the different icing algorithms including POD, impacted area, and volume using 40-km Eta data from a 24-h forecast valid at 0000 UTC 28 January 1994.

	POD_{All}	POD_{MOG}	Area ($\times 10^6$ km²)	Volume ($\times 10^6$ km³)
RAP	0.68	0.65	7.05	14.60
General	0.68	0.64	6.96	13.80
Unstable	0.24	0.24	6.69	5.79
Stratiform	0.53	0.53	4.94	5.76
Freezing rain	0.17	0.23	0.65	0.43
NAWAU	0.73	0.68	7.67	19.40
Category 1	0.73	0.68	7.66	19.30
Category 2	0.64	0.62	6.57	10.60
Air force	0.64	0.63	6.32	13.20
Cloud liquid water	0.41	0.41	3.43	2.86
NAWAU AIRMETs	0.79	0.75	4.68	18.91

Additionally, three quantities were utilized to evaluate overforecasting: POD_{No}, impacted areas, and impacted volumes. The POD_{No} is the probability of detection of negative pilot reports of icing and will not be addressed here but in B97 instead. The impacted area and volumetric coverage of model-generated icing predictions provide surrogate information about overforecasting. These two measures were used since it is inappropriate to use the more traditional false alarm ratio with the PIREP database (see B97). The impacted area values presented in subsequent tables are obtained by projecting a "shadow" of the icing at any level onto the surface of the earth and summing the area of each shadowed model grid box. Impacted volume is determined by simple summation of the grid volumes since each model grid point actually represents a volume. The impacted areas and volumes also screen out regions of the model domain that are over oceans where verification is nearly impossible due to a lack of PIREPs. Therefore, the maximum area coverage possible is that of the contiguous United States and immediate coastal regions, or approximately 9.5×10^6 km². The values provided in the following tables are computed using the above-described domain and not solely the region depicted in the preceding figures. Values for the CLW field are determined using a threshold minimum of 0.01 g kg⁻¹ to determine a binary yes/no for icing regions.

Tables 4 and 5 contain some statistical results for the 24-h Eta and 12-h MAPS icing forecasts presented in the previous section. The top values listed for each algorithm in Tables 4 and 5 originate from combining all individual classifications within each scheme to obtain a simple yes/no icing flag for the overall algorithm. The values listed beneath are obtained by running each sub-classification as a separate algorithm since each portion is not necessarily mutually exclusive from the others in the RAP or NAWAU algorithms.

Referring first to Table 4, one can see the RAP, NAWAU, and air force algorithms produced nearly the

TABLE 5. Same as in Table 4 except data formed from MAPS model using a 12-h forecast valid at 0000 UTC 28 January 1994.

	POD_{All}	POD_{MOG}	Area ($\times 10^6$ km²)	Volume ($\times 10^6$ km³)
RAP	0.55	0.50	5.91	11.98
General	0.54	0.49	5.65	10.63
Unstable	0.19	0.15	4.78	6.30
Stratiform	0.13	0.08	1.54	1.59
Freezing rain	0.13	0.17	0.72	0.36
NAWAU	0.68	0.66	7.09	16.90
Category 1	0.68	0.66	7.07	16.90
Category 2	0.42	0.36	4.73	7.63
Air force	0.42	0.37	4.83	10.94

same statistics for this case for all of the provided measures, except the NAWAU algorithm impacted a slightly larger area and volume while the air force algorithm impacted slightly less. Similarly, the PODs of the NAWAU algorithm were proportionately higher and the air force algorithm proportionately lower than the RAP algorithm. Throughout the WRIPEP period a trade-off was apparent in the RAP algorithm performance during sensitivity tests performed on the T and RH thresholds—decreases in area coverage nearly always lead to decreases in POD.

Statistical results for the NAWAU-issued AIRMETs are also shown in Table 4 and reflect similar results to the RAP and NAWAU algorithms except smaller area coverage and larger volume. Of course, the AIRMETs are not an independent product since they may be influenced by the icing predicted by the automated NAWAU algorithm. Furthermore, the AIRMETs are created with the help of other data sources (particularly satellite imagery and PIREP data), so it is not surprising that the AIRMETs have a similar POD and less area. Nonetheless, a comparison is warranted.

The explicit liquid water forecast by the Eta Model performed reasonably well by detecting 41% of the moderate and greater icing reports while only producing 3.4×10^6 km² area coverage (50% less area coverage than the RAP or NAWAU schemes). One additional item worth noting from Table 4 is the statistics shown for the stratiform portion of the RAP algorithm. This category appears to detect twice as many moderate and greater icing reports as the unstable category while impacting 25% less area. In fact, upon analyzing the entire WISP94 time period, this relatively good performance appears with high frequency (B97).

The similarities and differences between the three diagnostic algorithms are also apparent in the 12-h MAPS forecast statistics as shown in Table 5. Though both tables display statistics for the same valid time (0000 UTC 28 January 1994), direct comparisons should not be made for individual statistics since the MAPS data are from a 12-h forecast, whereas the Eta data are from a 24-h forecast. As mentioned in section 2, this is due to the fact that the MAPS model only forecasts 12 h in

length and there was no 1200 UTC model cycle for the Eta Model when the experimental version was operating in early 1994. The only conclusion to be made from comparing Tables 4 and 5 is that humidity predictions by the 12-h MAPS forecast must be less than humidity predictions from the 24-h Eta forecast. This is apparent from the statistics for the air force algorithm which has the most restrictive humidity thresholds as mentioned earlier.

6. Discussion

Unfortunately, the statistical analyses (referring to the entire WISP94 dataset, not solely the case study) do not always provide an accurate critique of the icing algorithm philosophy. Since the icing potential is completely dependent on model-predicted temperatures and relative humidities, the icing potential could be grossly incorrect due to model error. The statistical measures should instead be used as a baseline for comparison with other algorithms. The statistical analyses performed using the WISP94 data do not suggest that any one of the three automated algorithms (RAP, NAWAU, or air force) is superior to the others; however, the RAP scheme does provide some additional physical basis. Because of its classifications, the RAP algorithm may provide operational forecasters the opportunity to assess its skill and also refine the automated diagnostic output with additional data sources such as satellite data and surface observations.

To arrive at the threshold values for the RAP algorithm shown in Table 1, T and RH thresholds were adjusted in a series of steps. Through these steps, it became obvious that by restricting T and RH, thus decreasing area coverage, POD also decreased. The initial goal of fine-tuning T and RH to maintain a high value of POD while decreasing substantially the impacted area was not attainable. This realization led to the conclusion that utilizing thresholds, taken from Forbes et al. (1993), which detect most of the pilot-reported icing while impacting a reasonable area (as viewed by operational forecasters), might provide an optimal national-scale icing product.

On a practical level, in most cases it is easy to increase POD by increasing the forecast extent (i.e., area). If the forecaster or forecasting system is penalized for this increase by evaluating impacted area as a verification measure, then a trade-off is established between increasing area too much and adequately increasing POD. An algorithm that more efficiently detects icing PIREPs by impacting smaller areas will be rewarded by this approach, which seems like a reasonable goal. For example, the stratiform component of the RAP algorithm appears to detect the same proportion of PIREPs as the unstable component, but it forecasts a significantly smaller area (see Fig. 10 in B97). We believe this difference represents a true improvement in the capability of the stratiform component over the unstable compo-

nent, and it represents the meaningfulness of the area statistic as a verification measure. Furthermore, we feel that this POD/area approach requires forecast users to establish some minimum acceptable criteria for at least one of the verification statistics. For example, users could say, ''We must have a POD of 0.80 in order to use this product.'' Then the best forecasting scheme would be the one that attains a POD of at least 0.80 and has the smallest area coverage. Lastly, determining appropriate statistical criteria is the responsibility of the users and not the algorithm developers.

Statistically tuned diagnostic icing algorithms such as the ones presented here may contain nonphysical relationships as pointed out by Tremblay et al. (1996). All of the icing algorithms presented here have one nonphysical relationship: they all predict a flat or increasing frequency of icing at decreasing temperatures because they maintain or decrease relative humidity thresholds at lower temperatures. Observations by Tremblay et al. (1996) and statistical climatology studies of pilot-reported icing (Rasmussen et al. 1992) show that the number of icing reports decreases with decreasing temperatures. This is expected since at lower temperatures (particularly lower than $-15°$C) ice nucleation is more efficient [provided enough ice nuclei exist; Fletcher (1962)]. On the other hand, the RAP, NAWAU, and air force algorithms predict an increasing or level quantity of icing at lower temperatures (with the caveat that all algorithms cease predicting icing at approximately $-19°$ to $-22°$C). However, as discussed in the introduction, the algorithms presented in this paper are *attempting to infer cloudy environments in appropriate temperature ranges for icing.* They are simply trying to diagnose empirically where clouds might exist (on the basis of RH) and further refine the search to reasonable temperatures of icing observations (0 to $-20°$C). Obviously this is an oversimplification, but it also allows forecasters with access to operational numerical model's ''state-of-the-atmosphere'' variables to create a first-guess icing product.

Because of the discrepancy noted above, future work will focus most heavily on continued development–evaluation of explicit cloud liquid water schemes. Additionally, regional considerations in thresholding algorithms and incorporation of satellite imagery in icing diagnoses will be pursued. There is potential that the new *GOES-8/9* satellites will provide additional data to remove cloud-free regions from an automated icing analysis (Thompson et al. 1997; Lee and Clark 1995) as well as distinguish between cloud water and cloud ice. Furthermore, it is possible that by combining more than one channel of satellite data with additional remote sensing data, an excellent cloud liquid water *analysis* (hence aircraft icing *nowcast*) can be developed.

In summary, the icing algorithms presented in this paper are best used as first-guess icing potential fields. Because of the low RH thresholds used in the RAP-general and NAWAU-1 categories, their predictions will likely overforecast icing extent. Other data sources such as existing PIREPs, satellite data, and model-predicted cloud liquid water (if available) should be consulted to remove obvious icing-free regions. Known position errors or biases in numerical model forecasts should also be considered when viewing the icing products since the model predictions of T and RH are inputs to the icing algorithms. A forecaster may also consider eliminating regions with deep, continuous (in height) clouds with cold cloud tops (generally $T < -20°$C) since ice particles are likely to exist. However, multilevel clouds with dry air between represent a different and more difficult scenario. Other regions a forecaster should investigate closely are locations where the temperature/moisture sounding appears to be similar to the one in Fig. 1b. These soundings often represent a collision/coalescence process and ''nonclassical'' precipitation mechanisms (Isaac 1996). Any area of freezing precipitation including freezing drizzle, freezing rain, sleet, or ice pellets should be considered an immediate aircraft icing threat since supercooled large drops *must* exist somewhere within the vertical column above these reports (Bernstein 1996). Pilot reports of these weather conditions should be mandatory and would prove beneficial to the research community.

REFERENCES

Benjamin, S. G., K. A. Brewster, R. Brümmer, B. F. Jewett, T. W. Schlatter, T. L. Smith, and P. A. Stamus, 1991: An isentropic three-hourly data assimilation system using ACARS aircraft observations. *Mon. Wea. Rev.,* **119,** 888–906.

——, and Coauthors, 1994: The rapid update cycle at NMC. Preprints, *Tenth Conf. on Numerical Weather Prediction,* Portland, OR, Amer. Meteor. Soc., 566–568.

Bernstein, B., 1996: The stovepipe algorithm: Identifying locations where supercooled large drops are likely to exist. Preprints, *15th Conf. on Weather Analysis and Forecasting,* Norfolk, VA, Amer. Meteor. Soc., 5–8.

Brown, B. G., G. Thompson, R. T. Bruintjes, R. Bullock, and T. Kane, 1997: Intercomparison of in-flight icing algorithms. Part II: Statistical verification results. *Wea. Forecasting,* **12,** 890–914.

Cornell, D., C. Donahue, and C. Keith, 1995: A comparison of aircraft icing forecast models. Tech. Note 95/004, 33 pp. [Available from Air Force Combat Climatology Center, 859 Buchanan Street, Scott Air Force Base, IL, 62225-5116.]

Dudhia, J., 1993: A nonhydrostatic version of the Penn State/NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.,* **121,** 1493–1513.

Fletcher, N. H., 1962: *The Physics of Rainclouds.* Cambridge University Press, 386 pp.

Forbes, G. S., Y. Hu, B. G. Brown, B. C. Bernstein, and M. K. Politovich, 1993: Examination of conditions in the proximity of pilot reports of icing during STORM-FEST. Preprints, *Fifth Int. Conf. on Aviation Weather Systems,* Vienna, VA, Amer. Meteor. Soc., 282–286.

Huffman, G. J., and G. A. Norman Jr., 1988: The supercooled warm rain process and the specification of freezing precipitation. *Mon. Wea. Rev.,* **116,** 2172–2182.

Isaac, G. A., A. Korolev, J. W. Strapp, S. G. Cober, and A. Tremblay, 1996: Freezing drizzle formation mechanisms. *Proc. 12th Int. Conf. on Clouds and Precipitation,* Zurich, Switzerland, Int. Commission on Clouds and Precipitation, 11–14.

Knapp, D, 1992: Comparison of various icing analysis and forecasting techniques. Verification Rep., 5 pp. [Available from Air Force Global Weather Center, 106 Peacekeeper Dr., Offutt AFB, NE 68113-4039.]

Lee, T., and J. Clark, 1995: Aircraft icing products from satellite infrared data and model output. Preprints, *Sixth Int. Conf. on Aviation Weather Systems,* Dallas, TX, Amer. Meteor. Soc., 234–235.

Mesinger, F., Z. I. Janjic, S. Nickovic, D. Gavrilov, and D. G. Deaven, 1988: The step-mountain coordinate: Model description and performance for cases of Alpine lee cyclogenesis and for a case of Appalachian redevelopment. *Mon. Wea. Rev.,* **116,** 1493–1518.

NTSB, 1996: Aircraft accident report. Vol. 1. National Transportation Safety Board NTSB/AAR–96/01–PB96–910401, 322 pp. [Available from NTSB, 490 L'Enfant Plaza, S.W., Washington, DC 20594.]

Pobanz, B. M., J. D. Marwitz, and M. K. Politovich, 1994: Conditions associated with large-drop regions. *J. Appl. Meteor.,* **33,** 1366–1372.

Politovich, M., and B. Bernstein, 1995: Production and depletion of supercooled liquid water in a Colorado winter storm. *J. Appl. Meteor.,* **34,** 2631–2648.

Rasmussen, R., and Coauthors, 1992: Winter Icing and Storms Project (WISP). *Bull. Amer. Meteor. Soc.,* **73,** 951–974.

Schultz, P., and M. K. Politovich, 1992: Toward the improvement of aircraft icing forecasts for the continental United States. *Wea. Forecasting,* **7,** 492–500.

Schwartz, B., 1996: The quantitative use of PIREPs in developing aviation weather guidance products. *Wea. Forecasting,* **11,** 372–384.

Thompson, G., T. F. Lee, and R. Bullock, 1997: Using satellite data to reduce spatial extent of diagnosed icing. *Wea. Forecasting,* **12,** 185–190.

Tremblay, A., S. Cober, A Glazer, G. Isaac, and J. Mailhot, 1996: An intercomparison of mesoscale forecasts of aircraft icing using SSM/I retrievals. *Wea. Forecasting,* **11,** 66–77.

Vilcans, J., 1989: Climatological study to determine the impact of icing on low level windshear alert system. Rep. DOT–TSC–FAA–89–3, 330 pp. [Available from Volpe Transportation Systems Center, Cambridge, MA 02142.]

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.

Zhao, Q., F. H. Carr, and G. B. Lesins, 1997: A prognostic cloud scheme for operational NWP models. *Mon. Wea. Rev.,* **125,** 1931–1953.