Army Research Laboratory

An Evaluation of Three-Dimensional Weather Hazards Using Sounding Data and Model Output Data

by
Jeffrey E. Passner

Computational & Information Sciences Directorate
Battlefield Environment Division

ARL-TR-1046 **August 2000**

NOTICES

Disclaimers

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave Blank*)

2. REPORT DATE
August 2000

3. REPORT TYPE AND DATES COVERED
FINAL

4. TITLE AND SUBTITLE
An Evaluation of Three-Dimensional Weather Hazards Using Sounding Data and Model Output Data

5. FUNDING NUMBERS

6. AUTHOR(S)
J. E. Passner

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Research Laboratory
Computational & Information Sciences Directorate
Battlefield Environment Division
Attn: AMSRL-CI-EW
White Sands Missile Range, NM 88002-5501

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-1046

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783-1145

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARL-TR-1046

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT (Maximum 200 words)

The Atmospheric Sounding Program (ASP) is initialized by upper-air observations, either from standard rawinsonde observations (RAOB)s or output from a numerical model, the Battlescale Forecast Model (BFM). These data are decoded and processed before calculations are performed, giving the forecaster an overview of the atmospheric conditions at or near the ROAB launch site or the BFM grid point. The ASP uses these data to produce a series of weather-hazard products that can be used for analysis or forecasting to 24 h from the initial time of the BFM run. Included in these weather hazards are turbulence, icing, and clouds—three parameters, which have major influence in military operations.

14. SUBJECT TERMS
weather, hazards, BFM, soundings

15. NUMBER OF PAGES
59

16. PRICE CODE

17. SECURITY CLASSIFICATION OF THIS REPORT
UNCLASSIFIED

18. SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. SECURITY CLASSIFICATION OF ABSTRACT
UNCLASSIFIED

20. LIMITATION OF ABSTRACT
SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
prescribed by ANSI Std. Z39-18-298-102

Preface

The Atmospheric Sounding Program (ASP), developed by the U.S. Army Research Laboratory, is part of the U.S. Army Integrated Meteorological System Block II software. The program utilizes raw radiosonde data from the Automated Weather Distribution System as well as output data created by the Battlescale Forecast Model which runs on the Integrated Meteorological System. The ASP is employed operationally worldwide on the Integrated Meteorological System.

This report briefly principles how three-dimensional weather products are developed in the ASP and an evaluation of these products for operations.

Contents

Figures

Tables

Executive Summary

Introduction

The U.S. Army Research Laboratory, Battlefield Environment Division has developed the Atmospheric Sounding Program (ASP) to assist the Staff Weather Officer in making accurate weather predictions in the battlefield. The ASP uses data generated either by a mesoscale model, the Battlescale Forecast Model (BFM), or data from conventional soundings. The output of the ASP is a series of text packages or graphics that display weather products such as icing, turbulence, clouds, surface visibility, and thunderstorm probability.

Purpose

This report describes the data input and the different weather hazards that might interfere with military operations. However, the main emphasis of this report is on the evaluation of the derived three-dimensional weather products such as turbulence, icing, and clouds, by using both a sounding and output from the BFM.

Overview

The ASP is initialized by upper-air observations, either from standard rawinsonde observations (RAOB)s or output from a numerical model, the BFM. These data are decoded and processed before calculations are performed, giving the forecaster an overview of the atmospheric conditions at or near the RAOB launch site or BFM grid point. The ASP uses these data to produce a series of weather-hazard products that can be used for analysis or forecasting to 24 h from the initial time of the BFM run. Included in these weather-hazard products are thunderstorm probability, turbulence, icing, clouds, and surface visibility.

1.0 Introduction

The Integrated Meteorological System (IMETS) is a transportable, operational, automated weather data receiving, processing, and disseminating system utilized by U.S. Air Force weather forecasters in support of U.S. Army operations. The U.S. Army Research Laboratory (ARL) has formulated a number of weather products that will support the forecaster to make more precise and detailed weather decisions. Of most relevance to the military, is the impact of weather hazards on military operations. These hazards include three-dimensional (3-D) weather effects such as icing, turbulence, clouds, and two-dimensional variables such as surface visibility and thunderstorms. Earlier efforts at ARL have centered on applying sounding data to exhibit text and graphical output of these weather parameters. However, with the development and fielding of a mesoscale model, the Battlescale Forecast Model (BFM), short-term, weather forecasts (<=24 h) of these weather hazards are now being produced. [1,2]

The Atmospheric Sounding Program (ASP) is a program in the IMETS software environment that displays meteorological data for the forecaster as well as developing products that are placed into a gridded data base that can be accessed by users. The program ingests sounding data either from conventional radiosondes or from 3-D model output. Once these data are read, the program displays a skew T-log P diagram, and a weather hazards product, as well as textual information about other weather hazards.

This report is divided into the following sections, each with a different degree of detail.

- 1.0 - Introduction
- 2.0 - Data Input Methods into ASP
- 3.0 - Overview of the Weather Hazards
- 4.0 - Statistical Evaluation of the 3-D Weather Hazards
- 5.0 - Summary

2.0 Data Input Methods into ASP

2.1 Data Sources

Data for the ASP comes from two sources, either conventional radiosondes or gridded 3-D output from the BFM. Sounding data are delivered into the IMETS database by either the Automated Weather Distribution System (AWDS), the Automated Weather Network (AWN), or manual ingest. The BFM data are placed into a Gridded Meteorological Database (GMDB), where they can be accessed by the ASP. Prior to processing the data and performing the meteorological calculations, a quality control check of the upper-air sounding data is made. Some of the error checks include determining that all mandatory levels are available, a complete set of surface data is obtainable, and each level has both a pressure and height value. A check for consistency in the data (such as temperature and dew point values) is also conducted to ensure that the meteorological calculations can be completed without error.

Once the BFM data are accessed for a grid point, the ASP eliminates the surface-derived data or the first level. Surface temperature and moisture observations are normally recorded at 2 m above ground level (AGL); therefore, the ASP accepts the 2-m level as the surface level temperature and dew point. This eliminates many problems that would occur if the BFM zero-level data were used. The wind data at ground level is set to zero by the model and the model's 10-m level is used as surface wind. There is no other quality control of BFM data because the BFM program has its own quality control program. Reference 3 details the BFM quality control program. [3]

2.2 Sounding Data Using Rawinsondes

This method uses rawinsonde data in World Meteorological Organization format (*Federal Meteorological Handbook No. 4*). These data are commonly divided into different groups known as the TTAAs (mandatory levels), TTBBs (significant level temperatures), and PPBBs (significant level winds). The actual number of levels will vary depending upon the height attained by the balloon and the atmospheric structure. [4]

2.3 Sounding Data from BFM Model Output

Lee and Henmi describe the mesoscale domain as an area that can range from 2000 to 2 km. The U.S. Army is concerned with an area of 500 km or less, which ARL refers to as the "battlescale." With this scale in mind, ARL has adapted a hydrostatic model, Higher Order Turbulence Model for Atmospheric Circulation which has been modified for U.S. Army applications. [5,6]

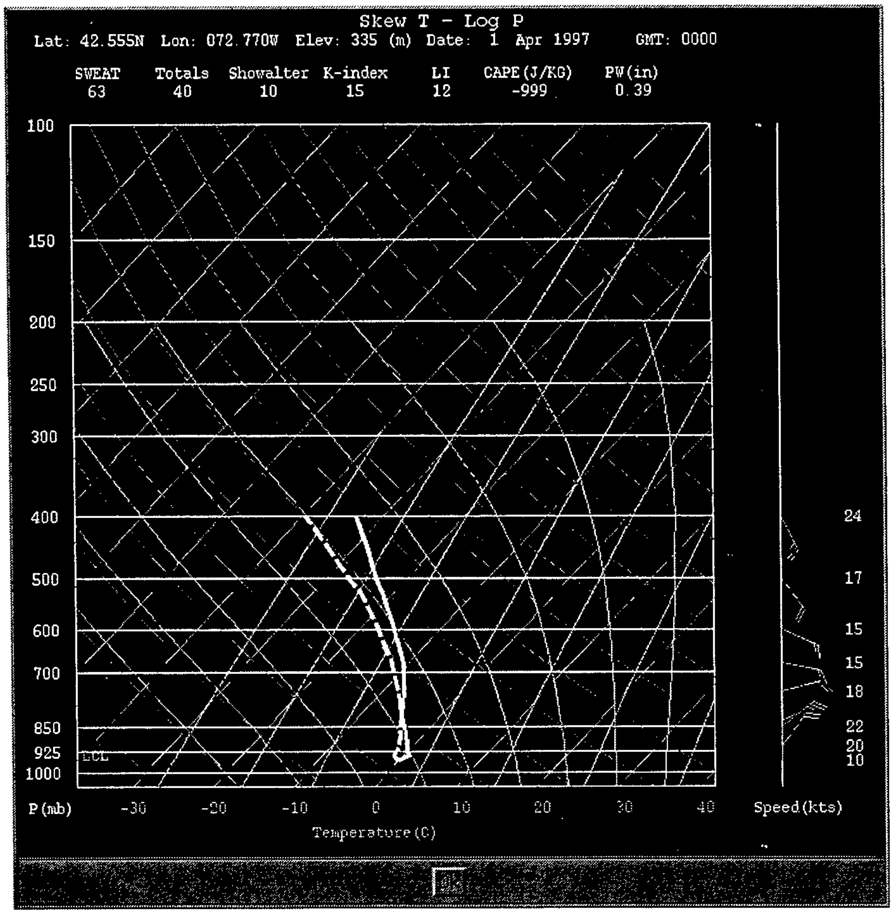
One of the main advantages of the BFM is that it takes into account local terrain features that assist in producing a fine-tuned forecast. By incorporating these features in a specific area, the forecaster does not need to be accustomed with the local terrain features and how it might influence nearby weather patterns. The BFM calculates intercepted solar radiant energy that can influence mesoscale wind fields. It uses the hydrostatic and quasi-Boussinesq approximations and has detailed surface boundary layer physics.

BFM initialization includes all observations from the area of interest, such as surface data, upper-air observations, and the 36-h forecasted Naval Operational Global Atmospheric Prediction System (NOGAPS) grid, which is issued by the U.S. Air Force Weather Agency (AFWA) via the U.S. Air Force AWDS. The NOGAPS grid points are spaced 1° latitudinal distance apart on the mandatory pressure surfaces, although much of the work done in this study was completed when NOGAPS was delivered with 2.5° data.

Lateral and time-dependent boundary conditions (large-scale forcing) are supplied from grid-point data close to the area of interest taken from NOGAPS output valid at analysis and forecast timest.

The BFM forecast is executed using these boundary conditions and area-of-interest raw data, as initialization guidance. The forecast solves towards forecast solutions dictated by global forecast gridded data, although boundary-layer mesoscale flows can be generated when the local terrain and radiation forcing dominate the large-scale dynamics (figure 1).

Figure 1. A 12-h forecast Skew T/Log P diagram from a BFM run.



The BFM-generated output for the grid includes the *u* and *v* horizontal wind vector component, potential temperature, and water vapor mixing ratio. These forecast fields are saved at 0,3,6,9,12,18, and 24 h from the base time of the model run; thus, it is possible to manipulate these data at various intervals over the forecast period.

3.0 Overview of the Weather Hazards

3.1 Weather Hazards Program

Today's weapons and sensors may be even more sensitive to weather than in the past. Performance of high-technology weapons such as the Advanced Tactical Missile System and the Apache helicopters can be degraded, as can many of the intelligence collection systems. The goal of the weather hazards program is to optimize weapon performance, assist in troop maneuvers, and aid the staff weather officer with weather guidance. [7]

The weather hazards program provides automated analysis and forecasts of what are considered "hazards" to U.S. Army operations. Additionally, many of the derived parameters in the ASP program are used by the Integrated Weather Effects Decision Aid (IWEDA) program. The IWEDA uses the ASP and BFM output to provide detailed information about why, when, and how weather impacts weapon systems (as well as their subsystems and components) and operations (figure 2). [8]

Often in weather forecasting, decisions must be made instantaneously; thus, it becomes beneficial to implement artificial intelligence techniques. The Weather hazards program, however, is not truly artificial intelligence because it uses statistical data, conventional computer programming techniques, and basic meteorological calculations as a "first guess" at the hazards. However, it becomes advantageous to use IF-THEN rules to assist in making weather products such as turbulence and clouds. The Weather hazards program makes an initial prediction and then gains information as it advances through the software in a top-down or forward-training methodology. This is called a Heterogenous Expert System because there is an integration of existing software with a Rule-Based Expert System. [1]

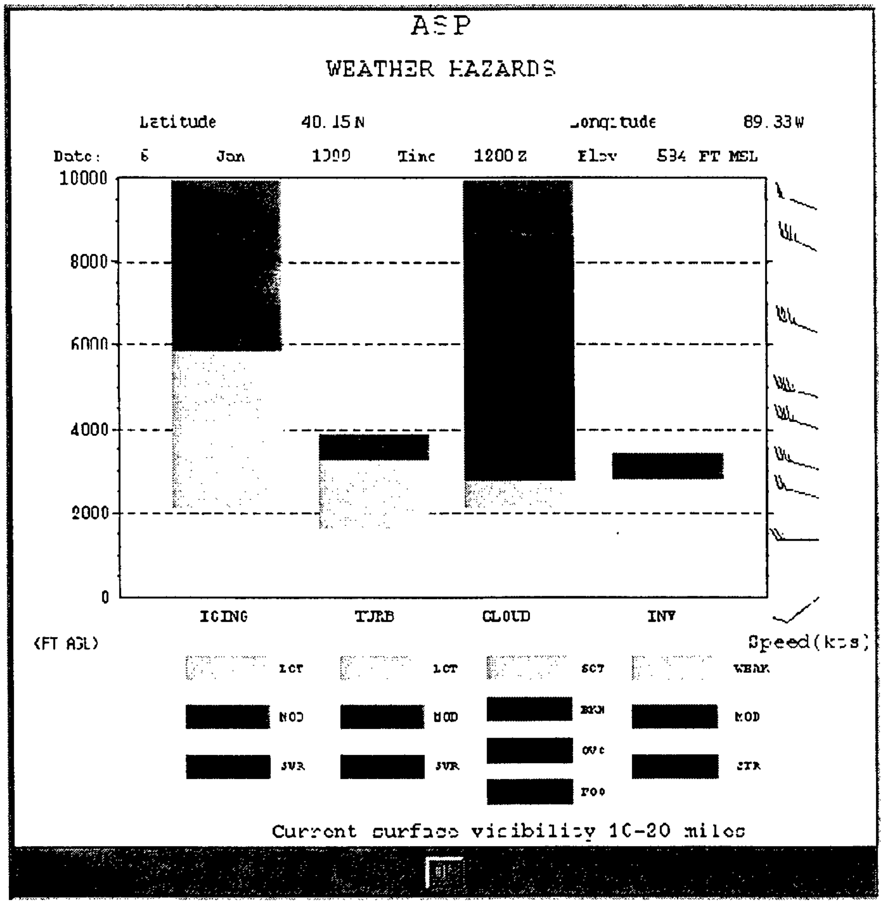


Figure 2. Plot of ASP weather hazards based on sounding observation.

3.2 Turbulence

Treating the atmosphere as a fluid, turbulence is generally a state of fluid in which there are irregular velocities and apparently random fluctuations. These oscillations in the atmosphere can adversely affect airframe performance and endanger U.S. Army aircraft. Turbulence is present in and near thunderstorms, as can be expected, based on dramatic updraft and downdraft speeds. Typically, a thunderstorm is a warning sign that turbulence will be present, and pilots need to make adjustments to their flight plans in the vicinity of convective clouds. [9]

Theoretical studies and empirical evidence have associated clear-air turbulence (CAT) with Kelvin-Helmholtz instabilities. Miles and Howard indicate that the developments of such instabilities require the existence of a critical Richardson number (RI) <=0.25. Keller notes that the RI is expressed as a ratio of the buoyancy resistance to energy available from the vertical shear. [10,11]

Of all the methods using a single sounding, the RI proved to make the most sense physically, since it included the influence of both the temperature and shear in the atmosphere. Additionally, in agreement with McCann, the RI displayed the most skill of several methods tested by using a single sounding. [12]

The U.S. Navy Fleet Numerical Meteorological and Oceanography Center uses the Panofsky Index (PI) to forecast low-level turbulence, where the low level is considered to be below 4000 ft AGL.

Based on analyses of raw RAOB data and corresponding pilot reports, early results of this work using the PI in the lower levels and the RI above 4000 ft, showed a strong bias to underforecasting turbulence. Thus, some additional rule checks were developed to ascertain that the derived numerical turbulence values made physical sense.

3.3 Icing

One of the most vital hazards forecast is aircraft icing not associated with convection. Generally, icing occurs at temperatures between 0 and -40 °C. Schultz and Politovich note that the accretion of ice on aircraft surfaces is controlled by two processes:

1. impaction of super-cooled cloud droplets on the aircraft and
2. the freezing of these droplets onto the airframe. [13]

In the ASP, three types of icing are considered:

1. rime,
2. clear, and
3. mixed.

Additionally, there are four icing intensities in the ASP:

1. *Trace icing*—the icing becomes noticeable on the aircraft.
2. *Light icing*—the accumulation of ice generates a problem for flights in excess of 1 h.
3. *Moderate icing*—the rate of accumulation presents a problem for short flights.
4. *Severe icing*—the rate of accumulation is so intense that deicing equipment fails to repress it.

Given the constraints of the single upper-air sounding as the data source, it was determined that the best approach to the analysis/forecasting of icing was to utilize the RAOB icing tool developed at the AFWA in 1980 (formerly, the U.S.

Air Force Global Weather Center). The RAOB technique uses the temperature, dew-point depression, and temperature lapse rate as a measure of instability of the layer.

The RAOB tool is essentially a "decision tree," in that it classifies icing by temperature, dew-point depression, and lapse rate. There are three main temperature groups:

1. -35 to -16 °C,
2. -16 to -8 °C, and
3. -8 to -1 °C.

These fundamental temperature classes are based on the theory of ice formation, with the first case, -35 to -16 °C, resulting in light rime icing in all cases. The second class, the -16 to -8 °C group, accounts for the mixed and rime cases, with the intensity based on the lapse rate or stability of the layer. The warmest group, the -8 to -1 °C group, is often the temperature range when clear icing is found; however, when the layer is stable, rime ice usually occurs. Clear ice most commonly occurs in layers of the atmosphere that are unstable and undergoing lifting.

In addition to the RAOB tool, a final case is added to account for severe clear icing. This situation occurs when there is a strong inversion about 100 mb above the surface; thus, the precipitation falls from a liquid state into a layer of subfreezing temperatures. This rapid change in temperature causes the relatively warm water droplets to spread quickly on the aircraft and cause clear icing to form.

A modification that has been applied in the ASP as compared to the original RAOB tool is an allowance for higher dew-point depressions, since the original RAOB tool was found to underforecast icing. Cornell's study showed that in the RAOB tool, dew-point depressions were too stringent, and his investigation of soundings showed that the mean dew-point depression for all icing types was 4.5 °C. This modification was made for the RAOB icing chart and is used currently in the ASP software. [14]

3.4 Clouds

Forecasting of cloud amounts, cloud heights, and cloud depth is essential for military operations. Clouds can degrade the effectiveness of many weapon systems by limiting flight paths, visibility, and making it impossible to identify targets and aircraft.

It was decided to approach the cloud-forecasting problem with empirical techniques, statistical data, and a rule-based IF-THEN set of code. This technique was best for using the single upper-air observation in the ASP. The main emphasis on the cloud program was to use relative humidity as the basis for cloud formation.

While it may appear that small differences in relative humidities are not consequential, work done by Walcek indicated that a two-to-three percent increase of the relative humidity could lead to a 15 percent increase in cloud cover. He also noticed, as in this study, that ceilings in the middle levels formed in lower humidity. Thus, a decision tree, or flow chart, is used to form the IF-THEN rules in the cloud program. [15]

Mesoscale models often display a dry bias. Schultz and Politovich observed that relative humidity values in excess of 55 percent between 500 to 1000 mb usually identify regions with widespread cloudiness on the Nested Grid Model. The BFM does not display such an extreme bias; however, clouds are often observed in layers with relative humidity well below values of saturation. In the ASP, a computer routine was developed for the formation of cumulus clouds. [13]

4.0 Statistical Evaluation of the 3-D Weather Hazards

4.1 Evaluation Results

The evaluation of the 3-D (vertical and horizontal) weather-hazard products has been ongoing for three years. The evaluation has resulted in program improvements; thus, this work has been evolving throughout the duration of the project.

A contingency table (table 1) provides a statistical method to display answers to binary YES/NO types of forecast evaluations. Some of the commonly used evaluation techniques include the probability of detection (POD), false alarm rate (FAR), the correct nonoccurrence (CNO), critical success index (CSI), true skill score (TSS), and bias. The calculations, based on the contingency table, are shown below.

Table 1. Contingency table for forecasted and observed weather event

	Forecasted YES	Forecasted NO
Observed YES	A	B
Observed NO	C	D

$$POD = \frac{A}{A + B} \tag{1}$$

$$FAR = \frac{C}{C + A} \tag{2}$$

$$CNO = \frac{D}{D + C} \tag{3}$$

Donaldson developed the CSI, and it was considered a standard in statistical evaluation, since it included three of the four elements in the contingency table. [16]

$$CSI = \frac{A}{A + B + C} \tag{4}$$

The CSI, however, does not take into account the null forecast (the D cell in the contingency table). Hanseen and Kuipers formulated an equation, which does factor in the null event, and it is called the TSS as seen in Eq. (5). This skill score is the ratio of observed skill to perfect skill and does not depend on the frequency of occurrence and nonoccurrence. [17]

$$TSS = \frac{(AD) - (BC)}{(A + B)(C + D)} \tag{5}$$

The bias in a forecast is the ratio of the number of positive forecasts to the number of observed events as shown in Eq. (6).

$$Bias = \frac{A + C}{A + B} \tag{6}$$

4.2 Turbulence Evaluation

Evaluation of clear-air turbulence is perhaps the most challenging of all the weather hazards. The most effective way to verify clear-air turbulence is to compare pilot reports with the forecasted values of turbulence for a layer and location in the atmosphere. Kane noted that pilot reports (PIREP)s contain the location, time, aircraft type, and turbulence severity and type. Naturally, there are variations in each PIREP, since these reports are the subjective judgement of the pilot and can vary from pilot to pilot. [18]

In the study by Kane, it was noted that 72 percent of pilots fail to specify the type of turbulence being encountered. Additionally, the reports are sporadic in time, location, and season with more PIREPs during the winter months, from sunrise to sunset, and closer to major airports. Kane also notes that a majority, 86 percent of PIREPs in the Continental United States, are reported from 1300 to 0100 UTC. Another obstacle in turbulence reporting is that it is nearly impossible to know the persistence of turbulence at a given level or location. Upon encountering turbulence, most pilots compensate by adjusting to another level—trying to avoid damage to aircraft, conserve time and fuel, and for the comfort for passengers. [18]

Another issue with turbulence verification in this study is that the PIREPs are compared to RAOB data. RAOBs are typically available only at scattered locations and two times a day (1200 and 0000 UTC). For evaluation purposes, it is assumed that a location 100 km from a ROAB site has analogous wind and temperature profiles as the RAOB site. This supposition makes it necessary to verify only those PIREPs close in location and time to the actual RAOB. An effort is made in this study not to accept any PIREPs more than 100 km from the airport and approximately three h from the time of the RAOB release. As an example, any PIREP before 1000 UTC and after 1400 UTC is not used for verification of 1200 UTC data because there are few flights before 1000 UTC, and the actual sounding time is closer to 1100 UTC than 1200 UTC; thus, the three-h time limit expires at 1400 UTC.

Using the BFM output, verification is limited to a one-h period surrounding the model forecast time. As an example, model forecasts of turbulence at 1800 UTC are compared to PIREPs from 1730 to 1830 UTC only.

A "correct" turbulence forecast was one within 100 km of the RAOB site and within two h of the RAOB release. Additionally, any PIREPs that included two intensities, such as LGT to MDT, were classified as the more extreme intensity (moderate in this example). Any turbulence forecast values near the height of the PIREP were accepted. For levels below 5000 ft AGL, the forecasted turbulence had to be within 1000 ft of the PIREP. From 5000 to 10000 ft AGL, the forecast had to be within 1500 ft of the PIREP, and above 10000 ft AGL, the forecast had to be within 2000 ft of the actual observed turbulence. As an example, a forecast for turbulence at 4000 ft AGL was only "correct" if a pilot reported turbulence between 3000 to 5000 ft AGL.

4.2.1 *PIREPs Statistics*

In this study there have been several statistical tests comparing turbulence forecasts derived from upper-air observations (the forecast) to PIREPs (the verification). The first study was in 1997, from February 25 to May 6, using 501 PIREPs. This study will be called the "1997 Study." The second study was conducted from November 1998 to April 1999 and is called the "1999 test." However, before comparing the forecasts against the observations, it is important to investigate the nature of PIREPs of turbulence. PIREPs of light, moderate, and severe turbulence were collected for both the 1997 and 1999 studies. It should be noted that reports of "neg," "smooth," or "no turbulence" mean that the pilot reported that there was no turbulence at the time of the report. It does not include any PIREPs where no turbulence report was submitted. Table 2 displays the results in the two different PIREP studies.

Table 2. PIREPs and intensity of the turbulence

	1997 Study	1999 Study
Samples	496	244
Report no turb	44%	39%
Report LGT turb	23%	27%
Report MDT turb	30%	30%
Report SVR turb	4%	4%

In the two samples, the number of "smooth"or "neg" turbulence events was 44 percent in 1997 and 39 percent in 1999. In a similar study, Marroquin investigated 17000 PIREPs and found 44 percent "no" turbulence cases and 56 percent "yes" reports of turbulence. As expected, severe or extreme turbulence is not reported often, most likely because it is rare, and pilots avoid flying conditions when severe or extreme turbulence is possible. [19]

Surprisingly, moderate turbulence is reported more often than light turbulence. There are probably two reasons for this.

1. As mentioned previously, all LGT to MDT reports were considered to be the more severe of the two (the moderate turbulence).
2. Pilots, especially in larger aircraft, probably do not report light turbulence because there is no harm to such airplanes and little discomfort for the passengers.

Another study was designed to look at the PIREPs by height. Table 3 shows a 1999 study of the number of PIREPs per 1000 feet by different layers of the atmosphere along with the percentage of light, moderate, severe, and no turbulence cases. A total of 242 PIREPs were used in compiling these results that indicate that the fewest PIREPs originate in the highest layers of the atmosphere and in the lowest layers, just above the surface.

Table 3. PIREPs per 1000 ft and turbulence intensity

Layer in ft (AGL)	PIREPs per 1000 ft	Percent no turbulence	Percent light turb	Percent MDT turb	Percent SVR turb
<= 2000	3.0	15	35	40	10
2000-4000	12.5	41	23	32	5
4000-8000	17.5	44	30	23	3
8000-14000	8.2	38	32	26	4
14000-20000	3.5	39	27	27	8
> 20000	2.5	33	24	41	2

Perhaps, this result is not surprising, since most pilots spend very little time in the lowest 2000 ft AGL and are often occupied with safely ascending or descending the aircraft. While these data in table 3 are limited, it appears that the most frequent layer for PIREPs is from the 4000 to 8000 ft AGL layer.

4.2.2 *Turbulence Evaluation for Sounding Sites*

The results of the analyses and forecasts made in the current study are displayed in this part of the report. Table 4 shows results from three different studies over the past three years, where the 1998 study was conducted from December 1997 to 16 April 1998, and the 1997 and 1999 were done simultaneously with the PIREP study in section 4.2.1. These three studies used the upper-air observations to calculate turbulence using the ASP computer routines.

Table 4. "YES/NO" turbulence statistics using upper-air observations

Turb statistics	1997 study	1998 study	1999 study	<5000 ft AGL 1999 study	5000-10000 ft 1999 study	>=10000 ft AGL 1999 study
Samples	501	100	298	36	86	104
POD	0.55	0.42	0.30	0.79	0.24	0.45
FAR	0.16	0.18	0.10	0.00	0.00	0.19
CNO	0.86	0.85	0.91	1.00	1.00	0.82
CSI	0.50	0.39	0.29	0.79	0.24	0.41
TSS	0.41	0.26	0.21	0.79	0.24	0.26
Bias	0.66	0.50	0.34	0.79	0.24	0.56

Table 4 shows the difficulty in forecasting and evaluating turbulence. Surprisingly, only slight changes have been made to the actual computer software to analyze and forecast turbulence over this three-year study. However, there is an obvious decrease in skill over the last three turbulence seasons. It is difficult to determine whether this is due to any of the software changes made or the testing method in this study. The tests each year were done at slightly different times of year, with the best results shown in the 1997 study that extended into early May. While it is not clear, it is possible that forecasting turbulence may be more challenging in the winter months. Thus, having some of the spring season in the 1997 study may have added some bias to the results. Additionally, the samples gathered over the three-year study were not standardized; the stations selected were random. Even with these uncertainties, these data cast very useful information about turbulence forecasts using the ASP.

The trend for correctly forecasting the nonevent is consistent through the entire study, along with a very low FAR. As might be expected with those statistical trends, there is a strong bias toward underforecasting the event (turbulence). The lower POD over the study is problematic. As mentioned, the techniques used, the PI below 4000 ft AGL, the RI above 4000 ft AGL, and simple rules based on temperatures and wind speed have not been dramatically changed; thus, the lower POD is most likely associated with the testing methods.

Based on the study in 1999, the third year of testing, the forecasts showed an intriguing trend with height. For predictions less than 5000 ft AGL, the POD was 0.79, and since there were no false alarms in the study of 36 samples, the

TSS was also 0.79. However, using the 86 samples between 5000 to 10000 ft AGL, the POD lowers significantly to 0.24 with a similar TSS. Above 10000 ft AGL, the POD increases to 0.45 with a TSS at 0.26.

It can be concluded from these data that either the PI is a very accurate tool in the lower levels, or turbulent motions are easier to forecast in these lower levels. This makes physical sense, since it is possible that most cold-season, low-level turbulence is the result of strong boundary-layer wind speed and directional shear—two elements that are derived easily from an upper-air observation. With an increase in height, more complex factors influence turbulence occurrence such as the sudden terrain differences, mountain waves, gravity waves, and vertical motions that are not derived easily from a simple upper-air observation.

The RI is one of two parameters used in this layer, along with a set of wind speed "rules." Most of the PIREPs used in the study were from the early morning hours in the cold season, so it can be expected that any mixing caused by solar radiation and heat transport is not a factor at this level. It is possible that turbulence formation is not well understood in this 5000 to 10000 ft AGL layer, or the error comes from the speed shear of the wind. However, the most likely cause for error in the 5000 to 10000 ft layer is that there were several reports of "OCNL LGT TURB" in this layer, which was considered an "incorrect" forecast. If these "LGT CHOP" PIREPs were considered as "insignificant" as they probably are for U.S. Army aircraft, the revised POD becomes 0.46 with a TSS of 0.46 for the layer—a much better skill score.

Still, the general trend of lower skill with height is not easily explained, but may be accounted for by more complex processes in the middle and upper atmosphere. Much more work and testing would need to be done along with much better data and a field study that could test these data against real-time pilot observations.

4.2.3 Turbulence Evaluation for the BFM Output

The turbulence routine for the BFM is not very different than the computer software used for sounding locations. The PI is employed as the main forecasting tool below 4000 ft AGL, while the RI is utilized above that level. The only difference is that the "rules" implemented to do consistency checks at the end of the program allow for turbulence to begin at lower wind speeds due to a slight BFM bias of underforecasting wind speeds. The BFM verification set also contains biases, since the model was run in poor weather conditions. There were two studies:

1. 1998 study: December 1997 through April 1998, which used 16 model runs and
2. 1999 study: 16 November 1998 through 16 April 1999, which used 15 model runs.

Most of these cases were initialized with 1400 UTC data and run for 24 h using 1° or 2.5° NOGAPS data, sounding data, and surface observations. All grids were 51*51 with 10-km spacing between the grid points using 16 vertical levels. PIREPs were used to verify the forecasts as closely as possible to every grid point. Table 5 shows the results of the model runs and turbulence verification.

Table 5. Turbulence and height statistics using BFM data

BFM turbulence statistics	**1998 study "YES/NO" forecast turbulence**	**1999 study "YES/NO" forecast turbulence**	**<5000 ft AGL turbulence (1999 study)**	**5000 to 10000 ft AGL turbulence (1999 study)**	**> 10000 ft AGL turbulence (1999 study)**
Samples	86	154	38	62	47
POD	0.58	0.50	0.68	0.24	0.64
FAR	0.18	0.09	0.10	0.00	0.05
CNO	0.78	0.91	0.85	1.00	0.95
CSI	0.49	0.48	0.63	0.24	0.62
TSS	0.36	0.40	0.52	0.24	0.58
BIAS	0.70	0.55	0.76	0.24	0.68

The turbulence results using these BFM data show nearly identical results and trends as using the sounding data (table 4) with a lower POD in the 1999 study than the previous year. A lower FAR and higher rate for correctly forecasting the nonevent lead to a slightly higher TSS. A surprising result is that using the BFM output data provides a better overall TSS (0.40 versus 0.21 for soundings in 1999 study). A possible explanation for this is that cases tested are often cases with "bad" weather, thus clear-cut cases of turbulence.

A second trend is the lower skill scores in the 5000 to 10000 ft range, with the same exact skill scores as the soundings (table 4). Again, is assumed that this is due to the large number of "OCNL LGT CHOP" reports in these levels that were considered incorrect in the statistics. The model output had lower skill scores below 5000 ft AGL but higher skill scores above 10000 ft AGL. It is uncertain as to why this occurs; however, it should be noted that the model top is generally lower than 25000 ft AGL, and because there are turbulence calculation problems near the top of the model, there was no effort to include any turbulence above 20000 ft AGL in these statistics.

While there were limited data for an evaluation of each forecast period, the results of the TSS are consistent over the 24-h forecast period. Since most of the model runs were started at 1400 UTC, the 18-h forecast often occurred in the middle of the night when pilots were not operating. Due to the limitations of these data, the "YES/NO" turbulence statistics will not be as complete as previous sets. There are no data available for the 18-h forecast in table 6.

Table 6. "YES/NO" turbulence forecasts by hours using BFM

Forecast hour	00 h	03 h	06 h	09 h	12 h	24 h
Samples	18	35	31	33	15	19
POD	0.55	0.67	0.43	0.47	0.36	0.62
FAR	0.17	0.13	0.00	0.08	0.00	0.11
TSS	0.44	0.52	0.43	0.38	0.36	0.45
Bias	0.67	0.70	0.43	0.52	0.36	0.69

The hourly results of the BFM turbulence forecasts show the same trends that exist with the soundings. The forecasts display a very low FAR and a strong bias to underforecast turbulence. This bias, is not surprising, given the difficulty in detecting and observing turbulence with PIREPs. Still, the trend indicates that much more work is required in this area to remove these biases through improving the rules or formulating a better understanding of the conditions that lead to turbulence. In the future, testing should be done to avoid any bias in areas with frequent and reliable PIREPs and numerous airports on the grid.

4.3 Icing Evaluation

Unlike turbulence, the basic "YES/NO" forecast of icing depends on the availability of moisture, formation of clouds, lapse rate, and temperatures below 0° C. Using an upper-air observation provides much of this information. However, the prime challenge for icing forecasts depends on these upper-air data to provide accurate information so that clouds can be forecasted. Once clouds have been predicted for a level, then the computer software activates the icing routine and can make projections of icing, icing type, and icing intensity. Many of the same difficulties encountered with turbulence-forecast evaluation exist with icing verification. Pilots tend not to report the "null" conditions, so it becomes difficult to know what percentage of flights are influenced by icing. Like turbulence, pilots often avoid weather situations where icing is expected.

The icing forecasts are evaluated in the same way as the turbulence forecasts. Only PIREPs within 100 km of the sounding site are used within two or three hours from the time of RAOB release. If a PIREP is "LGT-MDT," the moderate condition is considered the icing intensity. Any PIREPs of "trace" icing were grouped with "light" icing, while extreme icing was never observed in this study; thus, making "severe" icing as the most intense icing. For the BFM output, the verification period is again a one-hour period centered around the forecast hour. A model forecast for 2100 UTC is compared only to PIREPs between 2030 and 2130 UTC. All grids used were 51*51 with 10-km spacing between the grid points. Pilots report the location of the report in the PIREP; therefore, these locations were compared to the nearest grid point to the location of the icing event.

4.3.1 *Icing PIREPs*

Like turbulence, icing forecasts were compared to PIREPs, using random upper-air sounding locations around the United States at both 1200 and 0000 UTC. Icing was evaluated at all height levels, although most icing reports occur below 25000 ft AGL. In table 7 the icing types reported by the pilots are shown for the three study years.

Table 7. PIREPs of icing type

	1997 study	1998 study	1999 study
Samples	370	90	132
Nonevent	27%	33%	31%
Rime icing	75%	72%	67%
Mixed icing	18%	23%	23%
Clear icing	7%	5%	9%

As displayed in table 7, pilots do not often report the "null" or nonevent for icing. The icing types; rime, mixed, and clear icing are listed, assuming that a "YES" icing event has been reported. Rime icing is the most frequently reported by pilots, with little difference noted in each study year. The most infrequent condition is the clear icing, since clear icing usually occurs in unique conditions with near-zero temperatures in an atmospheric layer. Two additional studies were done to investigate the PIREPs for icing intensity, light, moderate, or severe. Table 8 shows the results of this study.

Table 8. PIREPs of icing intensity

Icing intensity	1997 study	1999 study
Samples	264	152
Nonevent	27%	24%
Light icing	65%	63%
Moderate icing	31%	33%
Severe icing	3%	3%

The above statistics indicate almost the exact same results for the two study years. It is interesting to note that the nonevent numbers are different than those in table 7. This is most likely because some pilots report icing type and not intensity or, conversely they report icing intensity and not the type. Another inquiry involved investigating the PIREPs of icing type and intensity with height. This study was done using data from the 1999 winter season. These results are shown in table 9.

Table 9. PIREPs of icing type and intensity by height

Icing type and intensity by height	<5000 ft (%)	5000-10000 ft (%)	10000-15000 ft (%)	>15000 ft (%)
Rime icing	76	64	78	76
Mixed icing	14	29	13	18
Clear icing	10	7	9	6
Light icing	70	62	53	68
Mod icing	26	36	44	32
Severe icing	4	2	3	0

In table 9 rime icing and light icing are the most commonly reported by the pilots, while clear icing and severe icing are the least likely to be reported. The trends of these data show a slight increase in mixed icing between 5000 to 10000 ft AGL, an indication that this layer may contain larger-sized, super-cooled water droplets. The "middle" atmosphere between 5000 to 15000 ft

AGL also shows an increase in icing intensity perhaps related to the lapse rate of this layer. In wintertime icing cases, the surface layer is often very stable with a steep inversion above the surface. However, just above this inversion layer there are often large layers of colder air aloft and steeper lapse rates that produce greater instability and more intense lifting of this air. Above 15000 ft AGL, the dominating ice type is rime and the intensity becomes lighter due to the colder temperatures and perhaps smaller droplet size.

4.3.2 *Icing Evaluation for the Upper-Air Observations*

By studying the trends in PIREPs it was possible to make minor changes to the original RAOB tool developed at AFWA. The RAOB tool (with the adjustments) was then used for the three-year study comparing PIREPs to proximity upper-air observations.

There have been three studies for the icing evaluations using upper-air soundings as the data source. These are for the same time periods as the turbulence studies.

1. 1997 study, 25 February to 6 May 1997;
2. 1998 study, December 1997 to 16 April 1998; and
3. 1999 study, 9 November 1998 to 1 April 1999.

Over the three-year study, the software for icing has not been changed significantly, although an error was found in the lapse-rate calculation and corrected for the third study (1999). Using the soundings, the statistical results are shown in table 10.

Table 10. "YES/NO" icing statistics using upper-air observations

	1997 study	1998 study	1999 study
Samples	264	92	153
POD	0.86	0.79	0.80
FAR	0.08	0.06	0.11
Nonevent	0.82	0.73	0.67
CSI	0.80	0.76	0.73
TSS	0.67	0.53	0.47
Bias	0.93	0.86	0.91

The data presented in table 10 show only slight differences from year to year using the sounding data. The only significant feature in the table is the decrease in the TSS during the 1999 study. This seems due to the decrease of forecasting the correct nonevent during that season, meaning that there was a trend to forecast a "YES" event when a "NO" event occurred. Because the non-event is included as part of the TSS, this led to a decrease in the TSS over the final year. Explanations for this can most likely be attributed to general improvements to the cloud-forecast program over the most recent years. Because the icing program depends on the existence of cloud layers, it appears that an increase in cloud layers also meant that icing forecasts increased, and there were more forecasts for a "YES" event. However, this increase was not "significant," since the overall bias of the icing forecast holds nearly steady over the three studies.

In general, these results give the user excellent guidance for "YES/NO" icing forecasts and are only slightly biased to underforecast the icing event. In table 11, the icing results are displayed as a function of height for only the 1999 study.

Table 11. "YES/NO" icing statistics by height using upper-air observations

	<5000 ft AGL	5000-10000 ft AGL	1000-15000 ft AGL	>= 15000 ft
Samples	24	68	35	25
POD	0.83	0.93	0.60	0.65
TSS	0.66	0.70	0.60	0.45
Bias	0.89	0.98	0.60	0.70

Table 11 shows the number of "forecasts" for each level and basic statistics. The trend is similar to the turbulence reports with the most PIREPs in the layer from 5000 to 10000 ft AGL. This level also displays the best skill using the U.S. Air Force RAOB tool and the in-house modifications by ARL. The skill scores are lower with increasing height in the atmosphere. Again, this may be related to the cloud-forecast program not detecting clouds in the middle atmosphere—a region where moisture is not as easily measured by soundings.

4.3.3 *Icing Evaluation for BFM Data*

Statistical evaluation for output data of the BFM are displayed in table 12.

Table 12. "YES/NO" icing statistics using BFM data

Icing statistics	1998 study	1999 study
Samples	78	112
POD	0.77	0.66
FAR	0.16	0.13
Non-event	0.43	0.59
CSI	0.67	0.61
TSS	0.21	0.27
Bias	0.92	0.76

In both years a relatively high POD was achieved along with a low FAR. However, the correct forecast of the nonevent was low. Note that the sample size was rather small (only seven nonevents in the sample in the 1998 study); thus, the TSS is significantly lower than the CSI in both years. There is a trend in the second season (1999) to underforecast the icing events. Compared to the skill scores using only upper-air data, the BFM "YES/NO" icing forecasts do deteriorate somewhat. This result can be expected because these data include icing forecasts to 24 h after the initial time of the model run. Additionally, the icing routine depends on forecasted clouds; therefore, the skill scores will be lower, since forecasting clouds is much more difficult using the vertical moisture profile with a model than with a sounding.

Table 13 displays icing forecasts for the forecast periods provided by the BFM output. Due to the small number of samples, only the POD is displayed in table 13. Additionally, there are not enough data for the 18-h forecast to be included in the table.

Table 13. "YES/NO" icing statistics by hours using BFM data

Hour	Samples	POD
00	18	0.73
03	20	0.73
06	25	0.50
09	12	0.64
12	15	0.71
24	14	0.54

In table 13, the probability of detecting an icing event is very low at the six-h forecast time. The initial forecast and the three-h forecast have the same skill, but by six h the model has a tendency to "lose" some of the initial moisture field as it begins to nudge to the NOGAPS data. By 24 h, the skill again decreases to only a 0.54 POD; however, given the small sample size, these results should be interpreted with caution.

Another trend in the model output and the upper-air data is the lower skill with increasing height (not shown). The FAR values remain low, but the POD decreases from 0.76 below 5000 ft AGL to 0.54 above 15000 ft AGL. In all cases, the bias is below 1.00; thus, the icing tool for the model underforecasts the icing event.

A major problem with the BFM RAOB tool is the inability to create any icing type other than rime icing. In the total sample of 66 cases, 94 percent of the forecasts were for rime icing, while 80 percent were observed to be rime. This is perhaps related to the limitations of a 16-level model missing many of the moisture layers and unstable layers that a sounding might capture. This would be a possible explanation for the small number of mixed and clear predictions of icing type.

A second difficulty with the RAOB tool for the BFM data was the intensity predictions. The BFM icing routine predicted trace/light icing 83 percent of the time and moderate icing in 17 percent of the cases compared to 63 percent and 33 percent observed by the pilots at the same time. There were no predictions of severe icing in the study.

4.4 Clouds Verification

The ASP calculates cloud amounts and heights from a variety of atmospheric variables such as relative humidity, season, time of day, location of station, and station elevation. If clouds are forecasted, the output includes the base of the cloud, the depth of cloud layer, the amount of cloud (scattered, broken, overcast), and the predominate type of cloud such as cumulus, stratus, and cirrus. In this study, the main emphasis was to evaluate the ceiling or the level where more than half the sky is covered by the cloud layer. There was no effort to assess parameters that were difficult to verify such as cloud depth or cloud type.

To evaluate the cloud amounts and heights, the ASP cloud forecasts were compared to Meteorological Aviation Routine Weather Reports which are coded weather observations at selected airports across the world. In the United States, many of these observations are taken by automated machines called the Automated Surface Observing System (ASOS) which do not report clouds above 12000 ft. To compensate for the growing number of ASOS stations, satellite photos were used in the 1999 study to account for the clouds analyzed or forecasted at the higher levels.

In the RAOB part of this cloud evaluation, only upper-air observations taken at 1200 and 0000 UTC were used. For the 1200 UTC upper-air observations, surface observations at 1100 UTC were used to evaluate the cloud forecasts, although observations from 1000 and 1200 UTC were often employed if weather conditions were changing rapidly. As an example, if the ASP forecast was for 8 OVC and the 1100 UTC observation reported CLR, a check of the record observations (SA) or special observations (SP) was done to ascertain that a ceiling did not form briefly and then dissipate before or after the 1100 UTC observation. Similarly, observations surrounding 2300 UTC were examined for rapidly fluctuating cloud heights or amounts when compared to the 0000 UTC upper-air observations.

For a forecast to be "correct," the height of the observed cloud had to be within:

- 1000 ft of the forecasted cloud height below 5000 ft AGL,
- 1500 ft between 5000 to 10000 ft AGL, and
- 2000 ft above 10000 ft AGL.

Additionally, the cloud forecasts were compared only to the surface observation at the upper-air sounding site location or observations within the same city area. For example, the Fort Worth, TX, sounding was evaluated against the Fort Worth, TX, observation. However, observations at the same hour were checked at the Dallas-Fort Worth airport and Love Field in Dallas, TX, to see if the general forecast verified across the entire city. A forecast for 15 BKN was accepted as "correct" even if Fort Worth, TX, reported clear skies, but Dallas-Fort Worth , TX reported 18 BKN. Since many National Weather Service offices have moved away from airports, the upper-air observations are often released a few miles from the larger airports in some cities. In these situations, more than one airport was used to verify the upper-air "forecast." Comments in the SA or SP were not used to verify a forecast. An observation of "clouds over mountains NW," was not accepted as verification for a cloud forecasted at the sounding site.

Finally, since the main emphasis of this study was on ceiling heights and amounts, scattered clouds were not considered "wrong" forecasts when there was no ceiling forecasted. However, if a ceiling was forecasted and only scattered clouds were observed, the forecast was considered wrong, although the difference between a scattered layer and a broken layer is often difficult to observe. When a broken layer was forecasted and an overcast layer was observed, the forecast was still correct as was a forecast for overcast conditions where broken clouds were reported.

Once an overcast layer was reported, there was no way to verify cloud layers or amounts above the overcast layer because the observer or ASOS could not see it. ASP's cloud routine continues to forecast and display these layers, but they could not be verified when a lower ceiling had already formed.

For the BFM output, the same verification system was used for the first two study periods; however, in the Winter 1999 study it was determined to permit a more liberal criteria on cloud height. The BFM is only a 16-layer model, of which four layers are within 30 m of the surface, where no clouds are permitted to form, and only fog is forecasted. BFM cloud forecasts were verified only at the hour of the observation, although SP observations within 30 min of the forecast was accepted for verification.

One major difference between the sounding and BFM forecasts should be noted. The ASP cloud program uses different relative humidity criteria for forecasting clouds when using the upper-air, short-term data and the BFM forecasts. For example, the upper-air data uses relative humidity values between 90 to 94 percent to form a ceiling between 4000 to 8000 ft AGL. However, the program forms a ceiling at the same level using BFM data with relative humidity values between 81 to 93 percent. These differences are based on a long-term study done by Passner. [2]

4.4.1 *Verification of Upper-Air Cloud Forecasts*

Using the listed criteria in the previous section, three different studies have been done for cloud verification. The forecasts were considered either "right" or "wrong" based on the cloud amount and cloud height within the time and space restraints already noted. These studies were done at 1200 and 0000 UTC during both the summer and winter seasons. The study periods were:

- Winter 1998, December 1997through 16 April 1998
- Summer 1998, 1 June - 11 through September 1998
- Winter 1999, 19 November 1998 through 29 March 1999

Table 14 shows results from the two "winter" studies and one summer study.

Table 14. Cloud statistics for all levels using upper-air observations

Hour and studies	Winter 1998 study	Summer 1998 study	Winter 1999 study
0000 UTC samples	105	181	261
0000 UTC correct	67%	79%	79%
0000 UTC wrong	33%	21%	21%
1200 UTC samples	325	331	203
1200 UTC correct	65%	66%	74%
1200 UTC wrong	35%	34%	26%

The results in table 14 indicate a general skill improvement over the two-year study; however, it should be noted that the final study, the one conducted in Winter 1999, was done at 16 specific sounding locations around the United States. These 16 sites were selected to represent a variety of different winter climates, while the Winter and Summer 1998 sounding sites selected were random and often tested in bad weather situations where it was known that the cloud forecast would be challenging. While it is difficult to explain the entire winter variation over the two seasons due to the more "standard" test, the computer software in the cloud routine program was also adjusted several times; so, perhaps the overall performance of the forecasts did improve. Another consideration would be the year-to-year and seasonal variety in weather patterns with one winter cloudier than another. The only examination of this factor was to see how many observations contained low clouds (ceilings <=3000 ft in the 1998 studies and <4000 ft in the 1999 study). Table 15 displays results from these investigations.

Table 15. Percentage of low-cloud ceilings in three studies

(%) observed low cloud ceiling	Winter 1998 (%)	Summer 1998 (%)	Winter 1999 (%)
0000 UTC	39	5	26
1200 UTC	54	23	32

In the first study, the Winter 1998 study, 54 percent of the 1200 UTC observations contained low clouds, while in the Winter 1999 study, with sites selected in a variety of locations, only 32 percent of the observations verified low clouds. At 0000 UTC in the first study (Winter 1998), 39 percent had low clouds, while in the third study (Winter 1999) 26 percent verified low clouds. Based on these results, it is likely that some of the improvement of the forecasts may be attributed to more clear days and fewer difficult cloud forecasts.

As mentioned above, in the Winter 1999 study, more standardization was added to the cloud verification by selecting upper-air locations evenly distributed through the United States and in different climate areas. Table 16 shows some of these cities and the percentage of correct forecasts for these stations.

Table 16. Cloud statistics for selected locations in winter 1999 study

Location	Samples	Percent correct
Long Island, NY	30	80
Tallahassee, FL	30	93
Miami, FL	31	81
Bismark, ND	31	77
Salt Lake City, UT	31	58
Denver, CO	27	81
Salem, OR	32	59
San Diego, CA	32	72

While the overall number of correct forecasts in the study was 77 percent, there is a wide variety of skill on the stations studied. The results include cloud forecasts for all levels at both 1200 and 0000 UTC. The highest skill level appears to be in the East and South, while the lower scores are noted in the Western region of the country, mainly in areas with difficult forecasting

Western region of the country, mainly in areas with difficult forecasting problems. Most interesting was the Salt Lake City, UT site where skill levels were low at both 1200 and 0000 UTC. This can be explained by the interaction of moisture from the Great Salt Lake which is close to the observing and sounding location. The rapid formation of snow showers and varying ceilings makes this site, located at a high elevation, a very difficult place to forecast clouds and other weather hazards.

Another challenge was at Salem, OR, located close to the Pacific Ocean and in direct line with rapidly moving bands of precipitation from frequent cold fronts and upper-air waves. Also, the slightly lower score at San Diego, CA, is due mainly to the 62 percent skill level at 1200 UTC. San Diego often experiences a marine layer in the morning, which is sometimes so shallow that the ASP routines do not forecast it correctly. The marine layer forms at varying times of the morning and often forms after the 1200 UTC sounding is released. Thus, this shallow, moist layer is not captured by the upper-air observation at the actual release time of 1030 UTC (0330 local time).

While much of this discussion has centered around the low clouds in the Winter 1999 study, it was possible to evaluate the higher clouds, since satellite photos were used to verify the high clouds. It is nearly impossible to determine the height AGL of the clouds using satellite, but it is possible to differentiate between high and low clouds using infrared photos. For simplicity, all cirrus clouds were assumed to be at 20000 ft AGL. Unfortunately, the sample size is smaller than ideal, but table 17 displays the cloud verification with height.

Table 17. Winter 1999 cloud statistics by height

Heights by hour, ceilings	<4000 number of samples	<4000 correct (%)	4000 to 8000 number of samples	4000 to 8000 correct (%)	8000 to 20000 number of samples	8000 to 20000 correct (%)
0000 UTC	54	76	18	50	12	8
1200 UTC	65	65	22	55	13	46

As noted in the winter study, most of the ceilings were low ceilings reported at 4000 ft AGL or less. Once a ceiling was formed, it was impossible to verify any clouds above that level, thus, resulting in very few ceiling samples at higher levels. The skill level does decrease dramatically with height, meaning that the ASP using an upper-air observation does not "forecast" ceilings above 8000 ft AGL with much skill. The most puzzling result is the 8 percent skill of the higher ceilings at 0000 UTC. The ASP computer routine is exactly the same for 1200 and 0000 UTC in the higher levels; therefore, it is uncertain why this occurs. Additionally, upper-air observations have difficulty recording moisture in the higher levels, especially during the winter months when the moisture is mainly in the form of ice at such levels. However, at the lower levels, the cloud routine does continue to be a very valuable guidance tool for the user, although the trend of lower skill at 1200 UTC is noticeable due to trouble with marine layers and morning stratus clouds.

4.4.2 *Verification of BFM Cloud Forecasts*

Three cloud-forecast studies were conducted using the BFM output; these three are similar to the ones done with upper-air observations. Because the BFM does not often reach over 25000 ft AGL, there was little effort to test for higher level clouds—especially cirrus clouds. With only 16 layers in the BFM, greater limits were allowed with respect to the height of the cloud layer in the Winter 1999 study. For the lower levels (below 10000 ft AGL) the cloud forecast was regarded as "correct" if the forecast was within 2000 ft AGL of the observation. Above 10000 ft AGL, the forecast was correct if it was within 3000 ft of the observed height. Thus, with this change, it might be expected that the Winter 1999 study might produce higher skill scores. Table 18 shows the evaluation of the BFM post-processed ASP cloud output.

Table 18. Cloud statistics for BFM by hours from model initialization time

Hour	Winter 1998 (%)	Summer 1998 (%)	Winter 1999 (%)
00	67 (52)	68 (43)	76 (51)
03	53 (15)	54 (33)	61 (54)
06	63 (43)	59 (31)	53 (57)
09	42 (26)	59 (33)	50 (52)
12	43 (43)	65 (32)	64 (45)
18	79 (19)	62 (47)	51 (41)
24	43 (51)	63 (40)	49 (43)

Note: Number of samples are in parenthesis

Noted that the BFM runs were predominately initialized between 1200 to 1400 UTC. The model was run at a variety of locations, with an effort to examine several different climate regions and varying weather situations. Some of the runs were completed in areas of clear weather and others were studied in difficult weather situations over complex terrain. The number of BFM runs was approximately 15 to 20 for each study.

The results in table 18 indicate that the cloud program performs with the most skill at the initial time of the model run, as might be expected. However, some differences are noted throughout the study. In the Winter 1998 study, the first study completed, there is a rapid decrease in the skill after the six-h forecast, a trend not as pronounced in the later studies. One of the reasons for this may be the general improvements in both the cloud program and the BFM software. Some changes in the nudging scheme may have helped to preserve the moisture from the initial hour to later hours before nudging completely to the larger-scale NOGAPS model in the later periods.

Surprisingly, the lowest overall skill is noted at the nine-h forecast time and not the later hours, such as 18 and 24 h after the model run began. A possible explanation is due to two complicated model factors. Because the model tends to nudge toward the scalar fields, such as temperature and moisture from the NOGAPS, a drier environment at those hours tends to reduce the relative humidity of each vertical layer of the model output. The ASP cloud program uses the relative humidity as a major parameter for the formation of clouds; thus, a lower relative humidity will reduce the cloud layers in the forecast clouds. Additionally, within the model itself, the lower relative humidity values will permit higher amounts of solar radiation to reach the surface grid field. This can lead to a higher surface temperature and even lower relative humidity values near the boundary, thus, reducing the chance of the ASP routine to forecast clouds.

It is interesting to note in table 18 that the cloud forecasts improve at the 12-h forecast period when the boundary layer begins to cool after sunset, especially in the winter season. This would reduce the radiation bias reaching the surface and would result in higher relative humidity values near the ground and more clouds.

Another trend seen on table 18 is the very consistent forecast skill during the summer months and more variable skill in the winter tests. These patterns agree with the upper-air studies, but may be due to the ability of the NOGAPS model to accurately forecast the moisture fields. In the winter, there are often strong synoptic-scale weather systems which bring frequent and rapid changes in many weather elements. The summer conditions are more likely to have convection and thunderstorm activity.

Even if the BFM lacks any type of cumulus parameterization routine, the ASP forecast cumulus clouds at peak heating hours during the summer months. The ASP does not create cumulus clouds in cooler environments, so in the winter storms, any cumulus field might be missed entirely in the forecast and could also help to explain the lower skill in the winter season.

A final look at the ASP model-derived cloud forecasts was to examine how accurate the cloud program is with height. To investigate, the atmosphere was divided into three layers:

- low clouds (<4000 ft),
- middle clouds (4000 to 8000 ft), and
- higher clouds (>8000 ft).

Table 19 displays the overall BFM cloud forecasts with height for all forecast hours.

Table 19. Cloud statistics for BFM by height

Cloud forecast by height (ft)	Samples	Number clouds low	Number clouds high	Percent correct	Cloud layer missed	Cloud layer forecast but not observed
<4000	246	32	16	64	72	17
4000 to 8000	63	11	6	49	28	4
>8000	34	3	3	32	18	5

In table 19, these data show that the skill of the cloud program decreases with height for the 1999 BFM runs. It also shows that the bias in the cloud routine is to forecast a ceiling lower than what was observed at the grid points. Additionally, much of the error is caused by "missing" a cloud layer (not forecasting it), rather than "overforecasting" a layer and not having one observed. In the lower levels, approximately 29 percent of all cloud layers were missed by the program. Overall, 34 percent of all clouds were missed and in only 8 percent of the cases was a layer forecasted and none observed.

Despite the high number of layers missed in the lower clouds levels, forecasting clouds in the lower levels is the strength of the cloud program. There is much concern, however, about the cloud forecasts in the middle and higher levels. Note that the most of the model runs conducted in the entire study used 2.5-NOGAPS data rather than 1° NOGAPS data now being used by the BFM. It is not certain that this improved resolution would enhance the cloud forecasts significantly, although it may provide better initial data and better moisture forecasts—especially in the early forecast periods.

The results displayed in table 19 (the BFM data) do not vary significantly from results in table 17 (the upper-air data). This is an indication that the cloud routine performs with about the same skill using the sounding data and BFM data. Table 20 shows a comparison between the 00-h BFM forecast and the sounding data at a variety of stations. The number of correct forecasts are almost identical using both methods.

Table 20. Comparison of upper-air observations and BFM initial-hour cloud analyses

	Winter 1998 % correct	Summer 1998 % correct	Winter 1999 % correct
BFM 00-h forecast	67	68	76
Upper-air obs	65	66	74

Results from the cloud forecasts will vary considerably with each different BFM run, mainly depending on the quality of initial data, NOGAPS forecast data, and how well the radiation parameters are handled. Also, the results seem best in cases where there is little variety in moisture fields, and the results seem most suspect in cases of rapidly changing weather and in regions of complex terrain.

5.0 Summary

The IMETS is an automated weather data receiving, processing, and disseminating system utilized by U.S. Air Force weather forecasters in support of U.S. Army operations. The ASP, a component of the IMETS software, calculates, interpolates, and displays meteorological data for the forecaster. The ASP uses sounding data either from 1200 or 0000 UTC upper-air observations or from BFM model output. The influence of 3-D weather hazards on tactical operations is of most concern to military leaders. These hazards include icing, turbulence, and cloud layers.

In the ASP, most applications used in the program are either flow chart type diagrams, expert system approaches using a set of rules, or regression equations designed for general and worldwide use. Turbulence is analyzed and forecasted in the ASP by using the PI below 4000 ft AGL and the RI above 4000 ft AGL. For icing, the RAOB tool originated at AFWA has been modified and is now used in the ASP.

Cloud forecasts were developed through careful investigation of moisture properties on skew-T diagrams through many different weather environments. This part of the ASP is the most "rule-based" in its design and uses a series of IF-THEN rules based on relative humidity, height of level, time of the day, season, and location of the station.

Over the past several years, detailed evaluation of these 3-D weather elements has demonstrated how effectively the products are, using both sounding data and output from the BFM. While it is vital to remember that weather forecasts of any type should still be in the hands of humans, the guidance provided by the BFM and the ASP post-processed parameters does assist the user in most military situations. All the forecasts, turbulence, icing, and clouds degrade with height using either data source, most likely due to the difficulty in measuring the atmosphere and forecasting complex interactions of atmospheric motions with more limited data. It is surprising that the tests in this study indicate that forecasting skills increase near the surface where there is more interaction between land, water, and air, but apparently better measurements from the sounding and higher vertical resolution of the BFM provide excellent skill scores of turbulence, icing, and clouds in the lower levels.

It can be concluded here that it is essential that meteorologists continue to resolve the vertical structure of the atmosphere with even more precision than exists today. There is no question that the more layers in a sounding and in a model the better the resulting forecasts can be. Additionally, there is a need for continued research into the mechanisms of turbulence, icing, and cloud formation. This obvious two-pronged approach, improving the observations and studying the motions of the atmosphere, can help to improve some of the challenging forecasting problems that influence military and nonmilitary aviation.

Still, given the current limitations of atmospheric measurements and difficult obstacles of weather forecasting, the results presented in this report do provide much confidence that the technology developed from the ASP and BFM provide optimal guidance in forecasting for U.S. Army operations. Work can be done to upgrade these forecasting tools and subsequent evaluations can be completed without the biases specified in these studies, but it is also integral to understand the constraints of the forecasting tools currently in use. With this knowledge, users can hopefully use these forecasts to the best of their ability.

References

1. Passner, J.E., "Expert Systems and Empirical Rules Used for Army Operations on IMETS," *Proceedings of the 13th Conference on Weather Analysis and Forecasting*, Vienna, VA, pp. 608-611, 1993. (Unclassified)

2. Passner, J.E., *The Atmospheric Sounding Program: An Analysis and Forecasting Tool for Weather Hazards on the Battlefield*, ARL-TR-1883, U.S. Army Research Laboratory, White Sands Missile Range, NM, 1999. (Unclassified)

3. Henmi T. and R.E. Dumais Jr., *Description of the Battlescale Forecast Model*, ARL-TR-1032, U.S. Army Research Laboratory, White Sands Missile Range, NM, 1998. (Unclassified)

4. U.S. Department of Commerce, *Federal Meteorological Handbook No. 4, Second Edition, Radiosonde Code*, Washington, D.C., 1976. (Unclassified)

5. Lee, M.E., "U.S. Army Battlescale Forecast Model Concept of Operation," *Proceedings of the 1994 Battlefield Atmospherics Conference*, White Sands Missile Range, NM, 1995. (Unclassified)

6. Henmi, T., et al., *Battlescale Forecast Model and its Evaluation Using White Sands Missile Range Meteorological Data*, ARL-TR-569, U.S. Army Research Laboratory, White Sands Missile Range, NM, 1995. (Unclassified)

7. Menoher, P.E. Jr., "Weather Support to Army Operations," *Proceedings of the 11th Annual EOSAEL/TWI Conference*, U.S. Army Sciences Laboratory, White Sands Missile Range, NM, pp. 3-51, 1990. (Unclassified)

8. Sauter, David P., "The Integrated Weather Effects Decision Aid (IWEDA): Status and Future Plans," *Proceedings of the 1996 Battlespace Atmospherics Conference*, Naval Command, Control, and Ocean Surveillance Center, San Diego, CA, pp. 7581, 1986. (Unclassified)

9. American Meteorological Society, *Glossary of Meteorology Fifth Edition*, Boston, MA, 1959. (Unclassified)

10. Miles, J.W. and L.N. Howard, "Notes on Heterogenous Shear Flow," *J. Fluid Mech*, **20**, pp. 331-336, 1964. (Unclassified)

11. Keller, J.L., "Clear Air Turbulence as a Response to Meso- and Synoptic-Scale Dynamic Processes," *Mon. Wea. Rev*, **118**, pp. 2228-2242, 1990. (Unclassified)

12. McCann, Donald W., "An Evaluation of Clear-Air Turbulence Indices," *Proceedings of 5th International Conference on Aviation Weather Systems,* American Meteorological Society, Vienna, VA, pp. 449-453, 1993. (Unclassified)

13. Schultz, P. and M.K. Polotovich, "Toward the Improvement of Aircraft-Icing Forecasts for the Continental United States," *Weather and Forecasting,* **7**, pp. 491-500, 1992. (Unclassified)

14. Cornell, D., C.A. Donahue, and K. Chan, *A Comparison of Aircraft Icing Forecast Models,* AFCCC/TN-95/004, U.S. Air Force Combat Climatological Center, Scott Air Force Base, IL, pp. 40, 1995. (Unclassified)

15. Walcek, Chris J., "Factors Influencing Regional-scale Cloud Cover: Investigations Using Satellite-Derived Cloud Cover and Standard Meteorological Observations," *Proceedings of the 4th Symposium on Global Change Studies,* American Meteorological Society, Anaheim, CA, pp. 235-236, 1993. (Unclassified)

16. Donaldson, R.J., R.M. Dyer, and M.J. Krauss, "An Objective Evaluator of Techniques for Predicting Severe Weather Events," *Preprints, 9th Conference Severe Local Storms,* Norman, OK, pp. 321-326, 1975. (Unclassified)

17. Woodcock, F., "The Evaluation of Yes/No Forecasts for Scientific and Administrative Purposes," *Mon. Wea. Rev,* 104 pp. 1209-1214, 1976. (Unclassified)

18. Kane, T, B.G. Brown, and R. Sharman, "A Preliminary Climatology of Upper Level Turbulence Reports," *Preprints of 8th Conference on Aviation, Range, and Aerospace Meteorology,* Dallas, TX, pp. 363-367, 1999. (Unclassified)

19. Marroquin, A., "Turbulence Forecasting Algorithms: Calibration, Comparison, and Verification," *Preprints of 8th Conference on Aviation, Range, and Aerospace Meteorology,* Dallas, TX, pp. 72-76, 1999. (Unclassified)

Acronyms

3-D	three dimensional
AFWA	U.S. Air Force Weather Agency
AGL	above ground level
ARL	U.S. Army Research Laboratory
ASOS	Automated Surface Observing System
ASP	Atmospheric Sounding Program
AWDS	Automated Weather Distribution System
AWN	Automated Weather Network
BFM	Battlescale Forecast Model
CAT	clear air turbulence
CNO	correct nonoccurrence
CSI	critical success index
FAR	false alarm rate
GMDB	Gridded Meteorological Database
IMETS	Integrated Meteorological System
IWEDA	Integrated Weather Effects Decision Aid
NOGAPS	Naval Operational Global Atmospheric Prediction System
PI	Panofsky Index
PIREPs	pilot reports
POD	probability of detection
RAOB	rawinsonde observation
RI	Richardson number

SA	record observation
SP	special observation
TSS	true skill score

Distribution

Copies

NASA MARSHALL SPACE FLT CTR 1
ATMOSPHERIC SCIENCES DIV
E501
ATTN DR FICHTL
HUNTSVILLE AL 35802

NASA SPACE FLT CTR 1
ATMOSPHERIC SCIENCES DIV
CODE ED 41 1
HUNTSVILLE AL 35812

US ARMY MISSILE CMND 1
AMSMI RD AC AD
ATTN DR PETERSON
REDSTONE ARSENAL AL 35898-5242

US ARMY MISSILE CMND 1
AMSMI RD AS SS
ATTN MR H F ANDERSON
REDSTONE ARSENAL AL 35898-5253

US ARMY MISSILE CMND 1
AMSMI RD AS SS
ATTN MR B WILLIAMS
REDSTONE ARSENAL AL 35898-5253

US ARMY MISSILE CMND 1
AMSMI RD DE SE
ATTN MR GORDON LILL JR
REDSTONE ARSENAL AL 35898-5245

US ARMY MISSILE CMND 1
REDSTONE SCI INFO CTR
AMSMI RD CS R DOC
REDSTONE ARSENAL AL 35898-5241

US ARMY MISSILE CMND 1
AMSMI
REDSTONE ARSENAL AL 35898-5253

PACIFIC MISSILE TEST CTR 1
GEOPHYSICS DIV
ATTN CODE 3250
POINT MUGU CA 93042-5000

ATMOSPHERIC PROPAGATION BRANCH 1
SPAWARSYSCEN SAN DIEGO D858
49170 PROPAGATION PATH
SAN DIEGO CA 92152-7385

METEOROLOGIST IN CHARGE 1
KWAJALEIN MISSILE RANGE
PO BOX 67
APO SAN FRANCISCO CA 96555

NCAR LIBRARY SERIALS 1
NATL CTR FOR ATMOS RSCH
PO BOX 3000
BOULDER CO 80307-3000

HEADQUARTERS DEPT OF ARMY 1
DAMI POI
ATTN LEE PAGE
WASHINGTON DC 20310-1067

DEAN RMD 1
ATTN DR GOMEZ
WASHINGTON DC 20314

US ARMY INFANTRY 1
ATSH CD CS OR
ATTN DR E DUTOIT
FT BENNING GA 30905-5090

HQ AFWA/DNX 1
106 PEACEKEEPER DR STE 2N3
OFFUTT AFB NE 68113-4039

PHILLIPS LABORATORY 1
PL LYP
ATTN MR CHISHOLM
HANSCOM AFB MA 01731-5000

PHILLIPS LABORATORY 1
PL LYP 3
HANSCOM AFB MA 01731-5000

AFRL/VSBL 1
29 RANDOLPH RD
HANSCOM AFB MA 01731

US ARMY MATERIEL SYST 1
ANALYSIS ACTIVITY
AMXSY
ATTN MR H COHEN
APG MD 21005-5071

US ARMY MATERIEL SYST 1
ANALYSIS ACTIVITY
AMXSY AT
ATTN MR CAMPBELL
APG MD 21005-5071

US ARMY MATERIEL SYST 1
ANALYSIS ACTIVITY
AMXSY CR
ATTN MR MARCHET
APG MD 21005-5071

ARL CHEMICAL BIOLOGY 1
NUC EFFECTS DIV
AMSRL SL CO
APG MD 21010-5423

US ARMY MATERIEL SYST 1
ANALYSIS ACTIVITY
AMSXY
APG MD 21005-5071

ARMY RESEARCH LABORATORY 1
AMSRL D
2800 POWDER MILL ROAD
ADELPHI MD 20783-1145

ARMY RESEARCH LABORATORY 1
AMSRL OP CI SD TL
2800 POWDER MILL ROAD
ADELPHI MD 20783-1145

ARMY RESEARCH LABORATORY 1
AMSRL CI LL
ADELPHI MD 20703-1197

ARMY RESEARCH LABORATORY 1
AMSRL SS SH
ATTN DR SZTANKAY
2800 POWDER MILL ROAD
ADELPHI MD 20783-1145

ARMY RESEARCH LABORATORY 1
AMSRL CI
ATTN J GANTT
2800 POWDER MILL ROAD
ADELPHI MD 20783-1197

ARMY RESEARCH LABORATORY 1
AMSRL
2800 POWDER MILL ROAD
ADELPHI MD 20783-1145

NATIONAL SECURITY AGCY W21 1
ATTN DR LONGBOTHUM
9800 SAVAGE ROAD
FT GEORGE G MEADE MD 20755-6000

US ARMY RSRC OFC 1
ATTN AMXRO GS DR BACH
PO BOX 12211
RTP NC 27009

DR JERRY DAVIS 1
NCSU
PO BOX 8208
RALEIGH NC 27650-8208

US ARMY CECRL 1
CECRL GP
ATTN DR DETSCH
HANOVER NH 03755-1290

US ARMY ARDEC 1
SMCAR IMI I BLDG 59
DOVER NJ 07806-5000

ARMY DUGWAY PROVING GRD 1
STEDP MT DA L 3
DUGWAY UT 84022-5000

ARMY DUGWAY PROVING GRD 1
STEDP MT M
ATTN MR BOWERS
DUGWAY UT 84022-5000

DEPT OF THE AIR FORCE 1
OL A 2D WEATHER SQUAD MAC
HOLLOMAN AFB NM 88330-5000

PL WE 1
KIRTLAND AFB NM 87118-6008

USAF ROME LAB TECH 1
CORRIDOR W STE 262 RL SUL
26 ELECTR PKWY BLD 106
GRIFFISS AFB NY 13441-4514

AFMC DOW 1
WRIGHT PATTERSON AFB OH 45433-5000

US ARMY FIELD ARTILLERY SCHOOL 1
ATSF TSM TA
FT SILL OK 73503-5600

US ARMY FOREIGN SCI TECH CTR 1
CM
220 7TH STREET NE
CHARLOTTESVILLE VA 22448-5000

NAVAL SURFACE WEAPONS CTR 1
CODE G63
DAHLGREN VA 22448-5000

US ARMY OEC 1
CSTE EFS
PARK CENTER IV
4501 FORD AVE
ALEXANDRIA VA 22302-1458

US ARMY CORPS OF ENGRS 1
ENGR TOPOGRAPHICS LAB
ETL GS LB
FT BELVOIR VA 22060

US ARMY TOPO ENGR CTR 1
CETEC ZC 1
FT BELVOIR VA 22060-5546

SCI AND TECHNOLOGY 1
101 RESEARCH DRIVE
HAMPTON VA 23666-1340

US ARMY NUCLEAR CML AGCY 1
MONA ZB BLDG 2073
SPRINGFIELD VA 22150-3198

USATRADOC 1
ATCD FA
FT MONROE VA 23651-5170

ATRC WSS R 1
WSMR NM 88002-5502

ARMY RESEARCH LABORATORY 1
AMSRL CI S
COMP & INFO SCI DIR
WSMR NM 88002-5501

ARMY RESEARCH LABORATORY 1
AMSRL CI E
COMP & INFO SCI DIR
WSMR NM 88002-5501

ARMY RESEARCH LABORATORY 1
AMSRL CI W
COMP & INFO SCI DIR
WSMR NM 88002-5501

DTIC 1
8725 JOHN J KINGMAN RD
STE 0944
FT BELVOIR VA 22060-6218

US ARMY MISSILE CMND 1
AMSMI
REDSTONE ARSENAL AL 35898-5243

US ARMY DUGWAY PROVING GRD 1
STEDP3
DUGWAY UT 84022-5000

USTRADOC 1
ATCD FA
FT MONROE VA 23651-5170

WSMR TECH LIBRARY BR STEWS IM IT WSMR NM 88002	1
US ARMY RESEARCH LAB AMSRL D DR D SMITH 2800 POWDER MILL RD ADELPHI MD 20783-1197	1
US ARMY CECOM INFORMATION & INTELLIGENCE WARFARE DIRECTORATE ATTN AMSEL RD IW IP FORT MONMOUTH NJ 07703-5211	1
ARMY RESEARCH LAB AMSRL CI EW J PASSNER WSMR NM 88002-5501	30
Record copy	1
TOTAL	92