Integrated Turbulence Forecasting Algorithm (ITFA): Quality Assessment Report

Aviation Weather Research Program
Quality Assessment Group

Barbara G. Brown[1], Jennifer L. Mahoney[2], Randy Bullock[1], Michael B. Chapmen[1], Chris Fischer[2,3], Tressa L. Fowler[1], Joan E. Hart[2,3], and Judy K. Henderson[2]

July 2002

[1] Research Applications Program, National Center for Atmospheric Research, Boulder, CO

[2] NOAA Research – Forecast Systems Laboratory, Boulder, CO

[3] Joint collaboration with the Cooperative Institute for Research in the Environmental Sciences (CIRES), University of Colorado/Forecast Systems Laboratory, Boulder, CO

Contents

Integrated Turbulence Forecasting Algorithm (ITFA): Quality Assessment Report

July 2002

Barbara G. Brown, Jennifer L. Mahoney, Randy Bullock, Michael B. Chapman, Chris Fischer, Tressa L. Fowler, Joan E. Hart, and Judy K. Henderson

Summary

This report summarizes assessments of the quality of forecasts of upper-level turbulence produced by the Integrated Turbulence Forecasting Algorithm (ITFA). ITFA was developed by the Turbulence Product Development Team of the Federal Aviation Administration's Aviation Weather Research Program (FAA/AWRP), and is currently being considered for transition to an operational product through the Aviation Weather Technology Transfer (AWTT) process.

The performance of ITFA forecasts has been evaluated over several winters by the AWRP Quality Assessment Group. Ongoing real-time and long-term evaluations are available on the Real-Time Verification System (RTVS; http://www-ad.fsl.noaa.gov/fvb/rtvs/turb/index.html), developed by the National Oceanic and Atmospheric Adminstration's Forecast Systems Laboratory (NOAA/FSL). In addition, in-depth analyses of the results have been undertaken at the Research Applications Program at the National Center for Atmospheric Research (NCAR/RAP). Both the real-time and post-analysis evaluations have involved meteorological/statistical verification of the turbulence forecasts. In addition, subjective evaluations of the accuracy of the forecasts have been provided by meteorologists at the National Centers for Environmental Prediction's Aviation Weather Center (NCEP/AWC) during the past three winters, and at Delta Airlines during winter 2001.

Because ITFA has evolved over the last several years, this report concentrates on results from the objective and subjective evaluations during winter 2002. These results are most relevant for operational use of ITFA. Trends and seasonal variations in the verification statistics over the past three years are also considered using results from RTVS.

The forecasts were verified using Yes and No turbulence observations from pilot reports (PIREPs) indicating either "moderate or greater" turbulence severity or "no turbulence". ITFA and several other turbulence algorithms were evaluated as Yes/No turbulence forecasts by applying a threshold to convert the output of each algorithm to a Yes or No value. A variety of thresholds were applied to each algorithm. The verification analyses were primarily based on the algorithms' ability to discriminate between Yes and No observations, as well as the extent of their coverage. In addition, forecasts based on Airmens' Meteorological Advisory (AIRMETs), the operational forecasts issued by the AWC, were evaluated to provide a standard of comparison. More than 1,000 individual ITFA forecasts were considered in this evaluation. The number of Yes (No) PIREPs considered in the evaluation ranged from 2,106 to 6,386 (861 to 2,680) depending on the forecast lead time.

Results of the evaluation indicate that ITFA is skillful at discriminating between Yes and No turbulence conditions. ITFA also provides relatively efficient forecasts, covering comparatively small volumes for a given turbulence detection rate. Using a threshold of 0.20, ITFA correctly classifies 68% of the Yes PIREPs and 69% of the No PIREPs, while covering approximately 23% of the airspace volume over the CONUS. The forecast quality is relatively insensitive to lead time, and is consistent through the atmosphere, above 20,000 ft. Detection rates vary from day-to-day, while volume coverage is relatively consistent from day to day. ITFA performance appears to be slightly better than the performance of other turbulence forecasting algorithms overall; however, the differences between ITFA and the "best" other algorithms are not statistically significant. Trends in ITFA performance over the last several years indicate that ITFA maintains its forecasting capability through the summer months, and that the overall skill in the forecasts has increased somewhat over time. The subjective evaluations indicate that the forecasters believe ITFA captures the turbulence well, but sometimes underestimates the extent and severity of a turbulence event.

The operational numerical weather prediction model that is used by ITFA [i.e., the Rapid Update Cycle (RUC)] evolved from a 40-km horizontal resolution to a 20-km resolution in mid-April 2002. Since the 20-km version is the new operational standard, it was important to evaluate changes in ITFA performance associated with the new model resolution. Results of a preliminary comparison indicate only a minor degradation in the verification statistics with the change to the finer-resolution model. This degradation may simply be attributed to the smaller area considered in matching the observations to the ITFA forecasts. Further tests, including analysis of additional cases and expanded PIREP matching procedures, are needed.

In summary, evaluations of ITFA over the last several years demonstrate that ITFA is a skillful forecasting algorithm that is generally able to discriminate between Yes and No turbulence PIREPs, with relatively efficient forecasts. The quality of ITFA forecasts is relatively insensitive to variations in the PIREPs used for the analyses and does not degrade with altitude. Long-term statistics indicate that ITFA also maintains its capability to correctly classify Yes and No turbulence situations throughout the year, including the summer months.

1. Introduction

This report summarizes basic results of an evaluation of the forecasting capability of the Integrated Turbulence Forecasting Algorithm (ITFA). This algorithm is under consideration for transition from experimental to operational through the Aviation Weather Technology Transfer (AWTT) process. ITFA was designed to predict clear-air turbulence (CAT) at altitudes above 20,000 ft over the continental U.S. (CONUS). It has been evaluated over several winter periods by the Quality Assessment Group (QAG) of the Federal Aviation Administration's Aviation Weather Research Program (FAA/AWRP) in specific algorithm intercomparison studies. In addition, long-term and real-time verification statistics on the performance of ITFA are available on the Real-Time Verification System (RTVS) developed by the National Oceanic and Atmospheric Administration's Forecast Systems Laboratory (NOAA/FSL) (Mahoney et al. 1997, 2002). The analyses in this report focus primarily on forecasts for winter 2002, although long-term performance trends are also considered. In addition to the real-time analyses, ITFA forecasts were evaluated in-depth in post-analysis.

Performance of ITFA forecasts has also been considered in several previous reports (Brown et al. 2000a,b,c, 2001; Mahoney et al. 2001b). In these studies, ITFA performance was compared to the performance of a large number of other turbulence forecasting algorithms. In most of the analyses included in this report, ITFA performance is compared to the forecasting performance of three other turbulence algorithms, as well as the operational turbulence forecasts. In addition to the objective evaluations, ITFA performance was evaluated during winter 2002 by forecasters at the NOAA National Centers for Environmental Prediction Aviation Weather Center (NOAA/NCEP/AWC). Basic results of this subjective evaluation of ITFA are also presented in this report.

The report is organized as follows. The study approach is presented in Section 2. Section 3 briefly describes the algorithms and forecasts that were included in the evaluation, and the data that were utilized are discussed in Section 4. The verification methods are described in Section 5. Results of the study are presented in Section 6. Finally, Section 7 includes the conclusions and discussion.

2. Approach

A total of 16 CAT algorithms were included in the winter 2002 RTVS evaluation of ITFA. Most of these algorithms also were included in previous evaluations. For post-analysis, and most of the results presented in this report, only four algorithms were considered. The algorithms were applied to data from the RUC-2 (Rapid Update Cycle, Version 2) model (Benjamin et al. 1998), with model output obtained from NCEP. Model forecasts issued at 1200, 1500, 1800, and 2100 UTC, with lead times of 3, 6, 9, and 12 hours and valid time between 1500 and 0000 UTC, were included in the post-analysis study. In addition, the turbulence Airmens' Meteorological Advisories (AIRMETs), which are the operational turbulence forecasts issued by

the AWC, were included for comparison purposes (i.e., this report is not intended as an evaluation of turbulence AIRMETs). Due to the emphasis placed on forecasting upper-level CAT, the evaluation focused on the region of the atmosphere above 20,000 ft, although in some cases results for forecasts above 15,000 ft are considered. For the RTVS analyses, forecasts issued between 1 January and 31 March 2002 are considered. The post-analysis includes results for 4 January 2002 through 15 April 2002.

In mid-April 2002 a new version of the RUC model became operational. A major difference between the old and new versions of the model is the increase in horizontal resolution from 40 to 20 km. Because the 20-km version of the model is the new standard, which will be employed by the operational version of ITFA, it is important to understand the sensitivity of ITFA performance to this change. This report includes a preliminary comparison of ITFA performance on the 20-km vs. 40-km versions of RUC for a short period in early April 2002.

The verification approach applied in the winter 2002 evaluation is identical to the approach taken in previous studies. In particular, the algorithm forecasts and AIRMETs were verified using Yes and No PIREPs of turbulence. The algorithm forecasts were transformed into Yes/No turbulence forecasts by determining if the algorithm output at each model grid point exceeded or was less than a pre-specified threshold. A variety of thresholds was utilized for each algorithm. The Yes/No forecasts were evaluated using standard verification techniques available for Yes/No forecasts, where observations are based on PIREPs. In addition, the amount of airspace impacted by the forecasts was considered. For most analyses, only PIREPs reporting moderate or greater (MOG) turbulence severity were included as Yes reports.

In evaluating an algorithm or forecast, it is important to compare the quality of forecasts to the quality of one or more standards of reference. Thus, the quality of the ITFA forecasts is compared to the quality of several other automated forecasting algorithms (e.g., Ellrod-1, DTF3; see Section 3), as well as to the quality of the operational forecasts (i.e., AIRMETs). However, it is important to emphasize that the algorithm forecasts and the AIRMETs are very different types of forecasts, with different objectives. ITFA forecasts generally are understood to be valid at a particular time. The AIRMETs, on the other hand, are valid over a 6-h period and are designed to capture turbulence conditions as they move through the AIRMET area over the period. Due to the differences between these forecasts, it is difficult to clearly compare their performance. However, in order to understand the quality of ITFA, it is necessary for comparisons between various forecasts to be made, and for ITFA forecasts to be compared to the operational standard, especially since both types of information will be available to users. The comparisons are made in such a way as to be as fair as possible to both the AIRMETs and ITFA, as described in Section 4, while still obtaining the information needed. Nevertheless, users of these statistics should keep these assumptions in mind when evaluating the strengths and weaknesses of each type of forecast.

A “forecaster evaluation” of algorithm performance was also included in the evaluation of ITFA. In this subjective evaluation, several forecasters at the AWC examined forecasts produced by ITFA and completed a questionnaire on a daily basis. The questionnaire concerned the synoptic meteorological conditions associated with observed turbulence events, as well as the

forecasters' perceptions of ITFA's performance. Results of this evaluation are considered only briefly in this report, and will be summarized more completely in a future report.

3. Algorithms and forecasts

The algorithms and forecasts that are considered in most of the analyses presented in this report are briefly described in this section. Further information about the algorithms and their development can be found in the references that are provided and in Sharman et al. (2002b); information about the algorithms included on RTVS is available through a link from the RTVS web site and in Sharman et al. (2002b). Operational forecasts of turbulence are also described.

DTF3: DTF3, developed by Marroquin (1995, 1998) is based on a simplification of the Stull (1988) TKE-γ model, which contains contributions from advection, diffusion, shear, convection and dissipation. Output from this algorithm is turbulent kinetic energy.

Ellrod-1: This index was derived from simplifications to the frontogenetic function. As such it depends mainly on the magnitudes of the potential temperature gradient, deformation and convergence (Ellrod and Knapp 1992).

ITFA : The ITFA forecasting technique uses fuzzy logic to integrate available turbulence observations (in the form of PIREPs) together with a suite of turbulence diagnostic algorithms (a superset of algorithms used in the verification exercise and others) to obtain the forecast (Sharman et al. 1999, 2000, 2002a,b). The suite of algorithms that is included is described in Sharman et al. (2002b). This algorithm was developed by the Turbulence Product Development Team of the AWRP. An example of an ITFA forecast is presented in Fig. 1. In this figure, the maximum ITFA values for a particular layer are shown, as well as the composite values based on the ITFA values in the whole column.

Richardson Number: Theory and observations have shown that at least in some situations patches of CAT are produced by what is known as Kelvin-Helmholtz (KH) instabilities. This occurs when the Richardson number (Ri), the ratio of the local static stability to the local shears, becomes small. Therefore, theoretically, regions of small Ri should be favored regions of turbulence (Drazin and Reid 1981; Dutton and Panofsky 1970; Kronebach 1964).

AIRMETs: AIRMETs are the operational forecasts of turbulence conditions. These forecasts are produced by AWC forecasters every six hours and are valid for up to six hours (NWS 1991). AIRMETs may be amended as needed between the standard issue times. The forecasts are in a textual form that can be decoded into latitude and longitude vertices, with tops and bottoms of the turbulence regions defined in terms of altitude. Unfortunately, some other more descriptive elements of the AIRMETs cannot be decoded and thus are not considered. For comparison with the forecasts from ITFA and other algorithms, the AIRMETs are evaluated over the same time window as the model-based algorithms.

4. Data

The data that were used in the evaluation include model output and PIREPs. Although lightning data were used in some previous evaluations to eliminate the effects of PIREPs related to convection (Brown et al. 2000a), it was determined in that study that this stratification had little impact on the results. Thus, lightning data are not considered in this study.

Model output was obtained from the RUC-2 model, which is run operationally at NOAA's NCEP, Environmental Modeling Center (Benjamin et al. 1998). The model vertical coordinate system is based on a hybrid isentropic-sigma vertical coordinate, and the horizontal grid spacing is approximately 40 km. The RUC-2 assimilates data from commercial aircraft, wind profilers, rawinsondes and dropsondes, surface reporting stations, and numerous other data sources. The model produces forecasts on an hourly basis; however, only the forecast and lead time combinations described in Section 2 were used in this study. Figure 2 depicts the RUC-2 domain and horizontal resolution. The verification analyses were limited to the domain covered by the AIRMETs, which also is shown in Fig. 2. Data for the 20-km version of the RUC model (which became operational on April 17, 2002) were obtained from the FSL mass store system.

Algorithms were applied to the model output files to create algorithm output files. This part of the process was undertaken by the ITFA algorithm developers. As part of this process, the algorithm output data were interpolated to flight levels (i.e., every 1,000 ft) rather than the raw model levels. The AIRMETs were decoded to extract the relevant location, altitude range, and other information.

All available Yes and No turbulence PIREPs were included in the study. These reports include information about the severity of turbulence encountered, which was used to categorize the reports. In particular, reports of moderate to extreme turbulence were included in the "Moderate-or-Greater" (MOG) category. Information about turbulence type (e.g., "Chop," "CAT") frequently is missing, and was ignored.

5. Methods

This section summarizes methods that were used to match forecasts and observations, as well as the various verification statistics that were computed to evaluate the ITFA and other forecasts.

5.1 Matching methods

The same methods were used to connect PIREPs to forecasts as in the previous evaluations (e.g., Brown et al. 2000a,b; Mahoney et al. 2001b). In particular, both the post-analysis and RTVS systems connect each PIREP to the forecasts at the nearest 8 grid points (four surrounding grid points; two levels vertically). However, the RTVS uses bi-linear interpolation to compute the appropriate forecast value, whereas the post-analysis system matches the PIREP to the most extreme (largest, except in the case of Richardson number) forecast value among the four surrounding gridpoints. As in previous evaluations, a time window of ±1 hour around the model valid time was used to evaluate both the algorithm forecasts and the AIRMETs.

5.2 Statistical verification methods

The statistical verification methods used to evaluate the results for winter 2002 are the same as the methods used in previous studies and are consistent with the approach described by Brown et al. (1997). More detail on the general concepts underlying verification of turbulence forecasts can be found in Brown and Mahoney (1998). These methods are briefly described here.

Turbulence forecasts and observations are treated here as dichotomous (i.e., Yes/No) values. AIRMETs essentially are dichotomous (i.e., a location is either inside or outside the defined AIRMET region). The algorithm forecasts are converted to a variety of Yes/No forecasts by application of various thresholds for the occurrence of turbulence. The thresholds used for ITFA, Richardson number, Ellrod-1 and DTF3 are listed in Table 1; thresholds for other algorithms included on RTVS can be found on the RTVS web pages. Thus, the basic verification approach makes use of the two-by-two contingency table (Table 2). In this table, the forecasts are represented by the rows, and the columns represent the observations. The entries in the table represent the joint distribution of forecasts and observations.

Table 1. Threshold values used to convert algorithm forecasts to Yes/No forecasts.

Algorithm	*RTVS*	*Post-analysis*
DTF3	0.2, 0.4, 0.6, 1.3, 2.0, 3.0	0.1, 0.2, 0.3, 0.4, 0.45, 0.5, 0.7, 0.9, 1.3, 2.0, 3.0
Ellrod-1	10^{-8}, $30x10^{-8}$, $40x10^{-8}$, $50x10^{-8}$, $70x10^{-8}$, $200x10^{-8}$	10^{-8}, $5x10^{-8}$, $10x10^{-8}$, $20x10^{-8}$, $25x10^{-8}$, $30x10^{-8}$, $35x10^{-8}$, $40x10^{-8}$, $50x10^{-8}$, $60x10^{-8}$, $70x10^{-8}$, $90x10^{-8}$, $120x10^{-8}$
ITFA	0.06, 0.08, 0.15, 0.2, 0.3, 0.4	0.02, 0.05, 0.07, 0.08, 0.09, 0.10, 0.13.0.17, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80,0.90
Richardson number	0.5, 1.0, 2.0, 4.0, 9.0, 15.0	0.5, 1.0, 2.0, 3.0, 4.0, 5.0, 7.0, 9.0, 12.0, 15.0

Table 3 lists the verification statistics used in this evaluation. As shown in this table, PODy and PODn are the primary verification statistics based on the 2x2 verification table. Together, PODy and PODn measure the ability of the forecasts to discriminate between (or correctly categorize) Yes and No turbulence observations. This discrimination ability is

summarized by the True Skill Statistic (TSS), which frequently is called the Hanssen-Kuipers discrimination statistic (Wilks 1995). Note that it is possible to obtain the same value of TSS for a variety of combinations of PODy and PODn. Thus, it always is important to consider both PODy and PODn, as well as TSS.

The relationship between PODy and 1-PODn for different algorithm thresholds is the basis for the verification approach known as "Signal Detection Theory" (SDT). For a given algorithm, this relationship can be represented by the curve joining the (1-PODn, PODy) points for different algorithm thresholds. The resulting curve is known as the "Relative Operating Characteristic" (ROC) curve in SDT. The area under this curve is a measure of overall forecast skill (e.g., Mason 1982), and provides another measure that can be compared among the algorithms.

Table 2 : Contingency table for evaluation of dichotomous (Yes/No) forecasts. Elements in the cells are the counts of forecast-observation pairs.

	Observation		
Forecast	***Yes***	***No***	***Total***
Yes	**YY**	**YN**	**YY+YN**
No	**NY**	**NN**	**NY+NN**
Total	**YY+NY**	**YN+NN**	**YY+YN+NY+NN**

It will be noted that Table 3 does not include the False Alarm Ratio (FAR), a statistic that is commonly computed from the 2x2 table. Due to the non-systematic nature of PIREPs, it is not appropriate to compute FAR using these observations. This conclusion, which also applies to statistics such as the Critical Success Index and Bias, is documented analytically and by example in Brown and Young (2000). In addition, due to characteristics of PIREPs and their limited numbers, other verification statistics (e.g., PODy and PODn) should not be interpreted in an absolute sense, but can be used in a comparative sense, for comparisons among algorithms and forecasts. Moreover, PODy and PODn should not be interpreted as probabilities, but rather as *proportions of PIREPs that are correctly forecast*.

Table 3: Verification statistics used in this study.

Statistic	***Definition***	***Description***	***Interpretation***	***Range***
PODy	YY/(YY+NY)	Probability of Detection	Proportion of Yes	0-1

Statistic	*Definition*	*Description*	*Interpretation*	*Range*
		of Yes observations	observations that were correctly forecasted	Best: 1 Worst: 0
PODn	NN/(YN+NN)	Probability of Detection of No observations	Proportion of No observations that were correctly forecasted	0-1 Best: 1 Worst: 0
TSS	PODy + PODn – 1	True Skill Statistic; Hanssen-Kuipers discrimination	Level of discrimination between Yes and No observations	-1 to 1 Best: 1 No skill: 0
Curve Area	Area under the curve relating PODy and 1-PODn	Area under the curve relating PODy and 1-PODn (i.e., the ROC curve)	Overall skill (related to discrimination between Yes and No observations)	0 to 1 Best: 1 No skill: 0.5
% Volume	[(Forecast Vol) / (Total Vol)] x 100	% of the total air space volume that is impacted by the forecast	% of the total air space volume that is impacted by the forecast	0-100 Smaller is better
Volume Efficiency (VE)	(PODy x 100) / % Volume	PODy (x 100) per unit % Volume	PODy relative to airspace coverage	0-infinity Larger is better

As shown in Table 3, two other variables are utilized for verification of the turbulence forecasts: % Volume and Volume Efficiency (VE). The % Volume statistic is the percent of the total possible airspace volume[4] that has a Yes forecast. VE considers PODy relative to the volume covered by the forecast, and can be thought of as the POD per unit volume. *The VE statistic must be used with some caution, however, and should not be used by itself as a measure of forecast quality*. For example, it sometimes is easy to obtain a large VE value when PODy is very small. An appropriate use of VE is to compare the efficiencies of forecasting systems with nearly equivalent values of PODy.

Use of these statistics is considered in somewhat greater detail in Brown et al. (2000a). In general, however, the argument presented in the previous paragraph can be extended to all of the

[4] The total possible area (limiting coverage to the area of the continental United States that can be included in AIRMETs) is 9.5 million km^2. Because the analyses are limited to 20,000 ft and above, the total possible volume thus is about 57 million km^3

statistics in Table 3; *none of the statistics should be considered in isolation* – all should be examined in combination with the others to obtain a complete picture of forecast quality.

As in previous turbulence forecast verification analyses, emphasis in this report will be placed on PODy, PODn, and % Volume. Use of this combination of statistics implies that the underlying goal of the algorithm development is to include most Yes PIREPs in the forecast "Yes turbulence" region, and most No PIREPs in the forecast "No turbulence" region (i.e., to increase PODy and PODn), while minimizing the extent of the forecast region, as represented by % Volume. ROC curve areas also will be considered as a measure of the overall skill of the forecasts at discriminating between Yes and No observations.

Quantification of the uncertainty in verification statistics is an important aspect of forecast verification that often is ignored. Confidence intervals provide a useful way of approaching this quantification. However, most standard confidence interval approaches require various distributional and independence assumptions, which generally are not satisfied by forecast verification data. As a result, the QAG has developed an alternative confidence interval method based on re-sampling statistics, which is appropriate for turbulence forecast verification data (Kane and Brown 2000). This approach is applied to some of the statistics considered in this report.

5.3 Stratifications

In some of the previous turbulence forecast evaluations, the verification results were stratified and limited using a variety of criteria applied to the PIREPs. These criteria included aircraft weight and proximity to lightning (Brown et al. 2000a). Results of the previous analyses indicated that stratifying by the aircraft weight and proximity to lightning criteria had little effect on the verification results for forecasts above 20,000 ft, except that it vastly reduced the number of PIREPs available for the analysis. Thus, these criteria were not applied in these analyses.

Most of the evaluations were limited to PIREPs and algorithm output above 20,000 ft. Two categories of reported severity are considered: (i) reports of any turbulence severity (light and greater) and (ii) reports of MOG severity. In most cases, results are presented only for MOG PIREPs. Generally, the results for All PIREPs are similar to those for MOG PIREPs, with somewhat smaller values of PODy. For most analyses, only forecasts at 20,000 ft and above were included. However, forecasts for 15-20,000 ft were considered in some cases. For most analyses, only "regular" PIREPs are included; additional reports received from United and Northwest Airlines are included in selected comparisons. In almost all cases the results are stratified by lead-time.

6. Results

Basic results of the winter 2002 verification analyses for ITFA and the other algorithms and forecasts are described in this section. The post-analysis verification analyses were limited to dates and times when algorithm output for all algorithms, as well as PIREP data and AIRMETs, were available, so all results would be comparable. A total of 382 3-h forecasts, 306 6-h forecasts, 228 9-h forecasts, and 151 12-h forecasts were included.

6.1 Overall results

Overall results for the 3-h lead time are shown in Fig. 3. The plots in Fig. 3 were created by combining the counts for all 3-h forecasts together. Figure 3a shows the relationship between PODy (MOG PIREPs) and 1-PODn, while Fig. 3b shows PODy versus % Volume. In these diagrams, the individual points on the algorithm curves represent individual thresholds used to create Yes/No forecasts. Results for better forecasts are located closer the upper left corners of the diagrams. The results in this figure indicate that all of the algorithms have similar skill at discriminating between Yes and No observations.

Figure 4 shows an example of results from RTVS, for 3-h forecasts issued at 1500 UTC. As noted earlier, the RTVS included a large number of algorithms in addition to those included in the post-analyses. The results shown in Fig. 4 are consistent (for the algorithms that are in common) with those presented in Fig. 3 for the post-analysis. Figure 4 also demonstrates that some of the algorithms that are included in the ITFA formulation (e.g., horizontal shear) have relatively little skill on their own; this result is consistent with previous studies (e.g., Brown et al. 2000a,b; Mahoney et al. 2001b).

The overall results can be examined in greater depth by selecting appropriate, comparable thresholds for each algorithm and comparing the individual statistics among the algorithms. As in previous studies, the rationale used for this process is to select thresholds that lead to a PODy value that is approximately the same as the value attained by the AIRMETs. Table 4 shows the results of this exercise for the 3-h forecasts. This table includes a variety of statistics associated with the specified thresholds. In addition, Table 4 includes (in the last column) estimates of the ROC areas (i.e., the areas under the curves in Fig. 3a). This statistic is not included for the AIRMETs since only one point is associated with the AIRMETs, which would lead to an unfair comparison.

Two values of PODy are included in Table 4 – one for All severities and one for MOG severities. In all cases, PODy (MOG) is somewhat larger than PODy (All). This result, which is consistent with previous results, suggests that the MOG PIREPs are somewhat easier for the forecasts to capture than are PIREPs associated with less severe conditions. The PODn values vary among the algorithms, with the largest value of PODn achieved by Richardson number [however, note that PODy(MOG) is somewhat smaller for Richardson number than for the other algorithms].

The TSS values in Table 4 provide a somewhat clearer comparison of the forecasting performance among the algorithms. Among the different forecasts and algorithms, the largest TSS values are achieved by the Richardson number, Ellrod-1, and ITFA. With regard to the ROC curve area, the best result is attained by ITFA. However, the ITFA ROC area is only slightly larger than the values achieved by the other algorithms, and the differences among the ROC areas are unlikely to be statistically significant.

In terms of the % Volume values in Table 4, the smallest (best) values are achieved by ITFA and Richardson number (again, this result is partially due to the somewhat smaller PODy for Richardson number). Because % Volume is strongly related to PODy, the small variations in

PODy in Table 4 may have had some impact on these results. Thus, in some cases it is more appropriate to consider the Volume Efficiency (VE) values. The best (largest) VE values in Table 4 were achieved by ITFA and the AIRMETs.

Table 4: Verification statistics for all 3-h forecasts (all issue times combined), for thresholds with PODy (MOG PIREPs) about the same as the PODy for AIRMETs.

Algorithm	Threshold	PODy (All)	PODy (MOG)	PODn	TSS	ROC Curve Area	Average % Vol	VE
AIRMETs	--	0.62	0.67	0.65	0.32	--	22.4	3.0
ITFA	0.17	0.69	0.74	0.60	0.34	0.75	29.2	2.5
	0.20	0.62	0.68	0.69	0.37		22.7	3.0
DTF3	0.90	0.61	0.66	0.67	0.33	0.72	24.5	2.7
Ellrod-1	0.0000004	0.64	0.69	0.66	0.35	0.73	24.8	2.8
Richardson	3.0	0.58	0.63	0.73	0.36	0.73	22.3	2.8

It is important to also consider variability in the verification statistics. Figure 5 shows curves of the 95% confidence intervals for the PODy values for ITFA, along with the curves for the other algorithms. Essentially all of the other algorithm curves are actually inside the pair of confidence interval curves, which indicates that the PODy values for ITFA are not significantly different from those for the other algorithms.

6.2 Comparisons among lead times

Figure 6 shows variations in the ROC and % Volume curves for ITFA for all of the different lead times considered (3, 6, 9, and 12 h), as well as the 0-h forecasts. The results in this figure indicate that the performance of the 0-h forecasts is somewhat better than the performance of the forecasts for the other lead times; this result is not surprising since the 0-h forecasts represent the fitted values of ITFA. In contrast, forecast performance does not appear to vary much among the other lead times: the curves for the 3, 6, 9, and 12-h forecasts appear to be quite close to one another. The results in Table 5 confirm that the ROC areas for ITFA change very little with lead time.

Table 5. ROC curve areas by lead time for all algorithms.

	Algorithm

Lead time (h)	ITFA	DTF3	Ellrod-1	Ruchardson number
3	0.75	0.72	0.73	0.73
6	0.74	0.71	0.73	0.72
9	0.74	0.71	0.74	0.71
12	0.73	0.71	0.71	0.71

Although the overall skill of ITFA forecasts does not change noticeably, as indicated by the curves in Fig. 6, it does appear that the location of individual points (i.e., representing particular thresholds) does vary with lead-time. This result is confirmed with the statistics presented in Table 6 for ITFA with a threshold of 0.2. In particular, PODy and % Volume decrease, and PODn increases with increasing lead time. This result suggests that the ITFA values are calibrated differently at different lead times. Thus, as lead-time increases, a smaller ITFA threshold is required to capture an equivalent proportion of YES PIREPs.

Table 6. Variations in verification statistics with lead time for ITFA forecasts with a threshold of 0.2.

Lead time (h)	PODy (All)	PODy (MOG)	PODn	TSS	Average % Vol	VE
3	0.62	0.68	0.69	0.37	22.7	3.0
6	0.57	0.63	0.72	0.35	20.5	3.1
9	0.54	0.59	0.76	0.35	18.6	3.2
12	0.52	0.57	0.76	0.33	17.2	3.3

Variations in the ROC and % Volume curves with lead time for the other algorithms are shown in Figs. 7 and 8, respectively. The ROC diagrams in Fig. 7 indicate little variation in this measure of skill with lead time. This result is confirmed by the ROC areas for the different algorithms shown in Table 5, which decrease only slightly with lead time for all of the algorithms. Figure 8 indicates that DTF3 and Richardson number are somewhat less skillful in terms of PODy vs. % Volume as lead times increase to 9 and 12 hours. Otherwise the statistics are fairly stable with lead-time.

6.3 Comparisons among PIREP types

For some analyses, additional PIREPs obtained directly from United (UAL) and Northwest Airlines (NWA) were included. The impact of these additional PIREPs on the verification results is illustrated in Fig. 9. Although there is some variation in the ROC diagrams with the addition of these reports, this variation is well within the confdence bounds for PODy, as illustrated in Fig. 5. The PODy vs. % Volume plots in Fig. 9 show very little variation with the PIREP type. These results are consistent among lead times and are similar to results presented previously (Brown et al. 2001).

6.4 Day-to-day variations

So far, only "bulk" verification statistics, accumulated over the entire verification period, have been presented. While these measures are relevant for evaluating overall performance, it also is important to consider day-to-day variations in the performance of ITFA. As an example, Fig. 10 shows RTVS plots of PODy vs. both % Volume and 1-PODn for 6-h ITFA forecasts issued at 1500 UTC, with points representing each forecast. These plots demonstrate the day-to-day scatter in the values. The time series plots of PODy and PODn values, also shown in Fig. 10, also demonstrate this day-to-day variation in the values of the verification statistics. These results are consistent with the confidence intervals presented in Fig. 5, and are similar for other issue/lead-time combinations.

Another way to examine day-to-day variations in the verification statistics is through box plots, which show the distributions of values of the statistics. As an example, Fig. 11 shows box plots of PODy and % Volume associated with individual ITFA thresholds, for 3-h ITFA forecasts. As shown in these plots, the distributions of PODy and % Volume decrease with increasing ITFA value. The PODy values are fairly variable (as indicated by the sizes of the boxes), especially for middle threshold values; this result is partly due to the fact that PODy is limited to the range 0-1, so is constrained to be less variable when approaching either 0 or 1. This variability is at least partly due to the small numbers of PIREPs that are available to verify any one forecast. The % Volume values exhibit less variability from day to day; in fact these distributions are quite narrow for any given ITFA threshold. Results for other lead times are consistent with the results shown in Fig. 11.

Figure 12 provides a closer look at the day-to-day variations in the statistics for two ITFA thresholds and for the AIRMETs. This figure suggests that day-to-day variability in the ITFA statistics is somewhat less than the variability in the AIRMET statistics; this results is particularly notable for the % Volume values. In general, except for % Area, the locations of the distributions are similar. For % Area, the ITFA distributions are much higher due to the fact that ITFA may produce isolated values that contribute to the total area.

6.5 Comparisons by altitude

During winter 2002, ITFA forecasts were produced down to an altitude of 15,000 ft. Figure 13 compares the overall results for ITFA and the other algorithms/forecasts for altitudes above 15,000 ft and above 20,000 ft. The curves in these figures indicate that there is little

variation in the overall results for the two altitude ranges, except for a slight improvement in ITFA performance relative to the other algorithms.

The height-series plots shown in Fig. 14 examine the variations in PODy and PODn with altitude in greater detail for the individual algorithms, for all of the 3-h forecasts combined. These plots, which are similar to those available on RTVS, are for specific thresholds for each algorithm: 0.2 for ITFA, 0.9 for DTF3, 0.0000004 for Ellrod-1, and 3 for Richardson number. In general, the results in Fig. 14 indicate that all of the algorithms have consistent capabilities at all altitudes.

Finally, the height-series plots in Fig. 15 consider variations in the ITFA height series with lead-time. As in Fig. 14, an ITFA threshold of 0.2 was used to create these plots. The results in Fig. 15 indicate that the variations of PODy and PODn with altitude are consistent across lead times.

6.6 RUC-20 comparisons

All of the analyses presented thus far, and in previous evaluations of ITFA, have been based on applying ITFA to the RUC-2 model, which had nominal 40-km horizontal resolution. On April 17, 2002, a new version of the RUC became operational, which has approximate horizontal resolution of 20 km. This version of RUC will be used by the operational version of ITFA, with forecasts produced on the 20-km grid. Thus, it is important to determine if changes in ITFA performance are associated with the change to this new version of the model. Fortunately, we were able to obtain 20-km RUC forecasts for a short period in April 2002 when the 40-km RUC was still operational. Thus, we have coincident model output for both versions of the model. The days included in the analysis are April 5-11, 2002, and the forecasts include 30 3-h forecasts, 24 6-h forecasts, 19 9-h forecasts, and 13 12-h forecasts. Only Ellrod-1 and ITFA were included in this evaluation.

The ROC curves for this comparison are presented in Fig. 16, and the % Volume curves are shown in Fig. 17. These curves are all much “bumpier” or less smooth than the curves that have been presented for the whole winter period; this characteristic is simply due to the small sample sizes considered in these analyses. In general – for both ITFA and Ellrod-1 – the ROC curves for the RUC-20 are located somewhat below the ROCs for the RUC-40 (Fig. 16); however, these differences are unlikely to be significant. Similarly, the % Volume curves for the forecasts based on RUC-20 are somewhat below the corresponding RUC-40 curves (Fig. 17).

This slight decrease in skill is not at all unexpected, when one considers the approach used to associate PIREPs with model output: that is, using the nearest four grid points. Four gridpoints represent a much larger area on the RUC-40 grid than on the RUC-20 grid, and thus allow a larger areal search for a Yes forecast than is allowed on the RUC-20 grid.

The results presented here indicate that the change in ITFA forecast skill associated with changing to the RUC-20 model is insignificant. Further analyses will be undertaken to determine the impact of increasing the number of gridpoints used to evaluate the 20-km version of ITFA; we anticipate that the slight decrease in skill noted here will disappear when the number of

gridpoints is increased slightly. In addition, a number of additional days in April 2002 will be included in future analyses.

6.7 Subjective evaluations

Over the last several winters, the QAG has sponsored subjective evaluations of the accuracy of ITFA and other turbulence algorithms. These studies have involved forecasters at the AWC and Delta Airlines, and dispatchers at ComAir. Results of previous studies have been described in several reports (Mahoney and Brown 2000; Mahoney et al. 2001a). During winter 2002, forecasters at AWC completed a large number of questionnaires regarding characteristics and locations of turbulence events and the quality of ITFA forecasts. Some basic results of this study are presented here; detailed results will be included in a report to be completed in the near future.

Figure 18 summarizes characteristics of the turbulence events that were identified by the AWC forecasters, including the location, cause, severity, and duration of the turbulence. Turbulence events in the Salt Lake City South region received the greatest percent of responses (Fig. 18a). This result may or may not indicate that more turbulence events occurred in that region. Other possible explanations for the large number of reports in this region include a larger volume of airspace and greater response rate from personnel monitoring turbulence in that region. The Boston region had the second greatest response rate. The fewest responses were obtained for the San Francisco South region. According to the AWC forecasters, more than half of the turbulence events were caused by the jet stream, while about a third were from "other" or unlisted causes (Fig. 18b). Mountain waves accounted for nearly 10% of the events. The few remaining turbulence events were caused by upper ridges, upper troughs, and convection. The maximum severity of the turbulence events, as classified by the AWC forecasters, was moderate or greater for over 90% of the cases (Fig. 18c). Finally, over a third of the turbulence events evaluated by the AWC forecasters had an unknown duration. However, nearly all of the turbulence events with a known duration exceeded 4 hours (Fig. 18d).

The AWC forecasters' perceptions of the quality of ITFA forecasts is considered in Fig. 19. The plots in this figure consider the overall ability of ITFA to capture the turbulence events, the appropriateness of the ITFA coverage, and the severity of the ITFA forecasts. The results in Fig. 19a indicate that nearly half of the ITFA forecasts captured the turbulence events well. About one third of the cases were judged by AWC forecasters to underforecast the turbulence. The remaining events, just less than 20%, were considered to be overforecasted by ITFA. The AWC forecasters also indicated that the coverage provided by ITFA for more than 50% of the events was too small (Fig. 19b), while coverage was too large (about right) for just over (under) 20% of the events. These results may relate to the magnitude of the ITFA threshold used to indicate areas of moderate-or-greater turbulence. Finally, Fig. 19c indicates that the severity of turbulence, as forecast by ITFA, was about right for roughly 40% of the events evaluated. For over half of the cases, the indicated severity was too light. Rarely, for less than 10% of the events, ITFA forecasted turbulence at an intensity that was too severe.

In summary, the AWC forecasters indicated that ITFA forecasts captured the turbulence events well, although the coverage provided by ITFA frequently was too small and the indicated

severity was too light. The latter two results may be related to the ITFA thresholds used by the forecasts; the objective verification results indicate that 3-h ITFA forecasts with a threshold of 0.3 (0.5) have PODy=0.47 (0.17) and PODn=0.86 (0.97). Further analysis is required to connect the particular responses to the type, location, and duration of the turbulence events. In addition, specific forecasts have been examined which are being considered in detail and compared to the responses provided on the questionnaire.

6.8 Long time series

It is instructive to consider long-term trends in the performance of ITFA and to examine variations in performance by season. Long-term statistics provided by RTVS are utilized for this analysis. Trends and seasonal variations in the AIRMET performance are also considerd simply to provide a baseline for the evaluation (i.e., these statistics are not presented here to provide an evaluation of the AIRMETs).

Monthly time series plots of the verification statistics for the AIRMETs are shown in Fig. 20, with the time series for ITFA presented in Fig. 21. The ITFA results consider a combination of all lead times; however results for individual lead times are consistent with those shown in Fig. 21. The AIRMET statistics appear to have a fairly regular seasonal cycle, with decreased PODy, increased PODn, and decreased TSS in the summer months, and the opposite effects in the winter months. This characteristic of the AIRMET statistics is most likely due to the fact that most turbulence conditions during the sumer are associated with convection, which is accounted for in the Convective SIGMETs issued by the AWC. Thus, fewer turbulence AIRMETs are issued during the summer months than during other times of the year. In contrast, the ITFA statistics show a much smaller seasonal variation (Fig. 21). In addition, the ITFA PODy values generally have had an increasing trend since January 2000, while the PODn values have had a more moderate decreasing trend. Thus, the overall trend in TSS for ITFA is somewhat increasing. Although this analysis considers all lead times combined, results for individual lead times are consistent with those presented in Fig. 21.

7. Conclusions and discussion

This report has summarized an evaluation of the upper level turbulence forecasts produced by ITFA. This exercise has followed several previous intercomparisons of forecasts produced by ITFA and other turbulence algorithms. The results obtained in winter 2002 are consistent with those obtained in previous exercises and suggest that ITFA is a potentially useful turbulence forecast product. In particular:

- ITFA forecasts are skillful, as measured by their ability to discriminate between Yes and No PIREPs of turbulence.

- ITFA forecast skill is similar to the skill of forecasts produced by a few other algorithms, such as DTF3 and Ellrod index, and the AIRMETs; ITFA skill is greater than the skill of many other algorithms (e.g., horizontal shear).

- Day-to-day variations in PODy can be fairly large (similar to variations in PODy associated with other algorithms and somewhat smaller than variations in the AIRMET PODy values), partly due to the small numbers of PIREPs available to verify a single forecast. Variations in the volume of airspace covered by ITFA are quite small.

- The skill of ITFA forecasts is relatively consistent throughout the year, with relatively small degradations in the summer months. ITFA does not show strong variations in performance with season.

- ITFA performance has followed a generally increasing trend over the last two years; it is unclear how much of this trend has been associated with variations in weather patterns over this period.

- ITFA forecasts perform consistently at all altitudes at 20,000 ft and above.

- AWC forecasters indicated that ITFA captured turbulence well, but for some turbulence events the extent of the forecast was too small and the indicated severity was too light.

- Initial results of evaluations of ITFA on the 20-km RUC model indicate a slight degradation in skill when the same verification approach is used as for the 40-km RUC – this degradation is expected to disappear when the verification methods are adjusted to be more appropriate for the 20-km RUC; further testing and analysis of additional cases are required.

The results described in this report are a small fraction of the verification results that are available. For example, a wide variety of verification information for ITFA, other algorithms, and the AIRMETs is available at the RTVS web site (http://www-ad.fsl.noaa.gov/fvb/rtvs/turb/index.html).

One additional strength of the ITFA approach, not identified above, is the algorithm's *adaptability* as new approaches are developed for forecasting turbulence and for combining

indices. For example, initial verification analyses applied to a statistical approach for combining indices in ITFA shows promise in improving detection rates and reducing forecast volumes (Brown et al. 2001; Tebaldi et al. 2002).

Acknowledgments

This research is in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy and position of the U.S. Government.

We would like to thank the members of the Turbulence Product Development Team for their support of the independent verification effort over the last several years. We also thank Jamie Wolff for making the algorithm output available during the real-time portion of the project and for the on-going re-computation of some of the fields. In addition, we would like to thank Huming Han of the FSL ITS division for obtaining the 20-km RUC data that were used in the evaluations; Jamie Braid (NCAR) for obtaining much of the other data that was analyzed and providing consistent support for the analyses; and Mike Kay (FSL) for assisting in collection and analysis of the subjective evaluation data.

References

Benjamin, S.G., J.M. Brown, K.J. Brundage, B.E. Schwartz, T.G. Smirnova, and T.L. Smith, 1998: The operational RUC-2. *Preprints, 16th Conference on Weather Analysis and Forecasting*, Phoenix, AZ, American Meteorological Society (Boston), 249-252.

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. *Weather and Forecasting*, **12**, 890-914.

Brown, B.G. and J.L. Mahoney, 1998: Verification of Turbulence Algorithms. Report, Available from B.G. Brown, NCAR, PO Box 3000 Boulder CO 80307-3000, 9 pp.

Brown, B.G., and G.S. Young, 2000: Verification of icing and turbulence forecasts: Why some verification statistics can't be computed using PIREPs. *Preprints, 9th Conference on Aviation, Range, and Aerospace Meteorology*, Orlando, FL, 11-15 Sept., American Meteorological Society (Boston), 393-398.

Brown, B.G., J.L. Mahoney, R. Bullock, J. Henderson, and T.L. Fowler, 2000a: Turbulence Algorithm Intercomparison: 1998-99 Initial Results. NOAA Technical Memorandum OAR FSL-25, 64 pp.

Brown, B.G., J.L. Mahoney, R. Bullock, T.L. Fowler, J. Hart, J. Henderson, and A. Loughe, 2000b: Turbulence Algorithm Intercomparison: Winter 2000 Results. NOAA Technical Memorandum OAR FSL-26, 62 pp.

Brown, B.G., J.L. Mahoney, J. Henderson, T.L. Kane, R. Bullock, and J.E. Hart, 2000c: The turbulence algorithm intercomparison exercise: Statistical verification results. *Preprints, 9th Conference on Aviation, Range, and Aerospace Meteorology*, Orlando, FL, 11-15 Sept., American Meteorological Society (Boston), 466-471.

Brown, B.G., T.L. Fowler, R. Bullock, and J.L. Mahoney, 2001: Turbulence algorithm evaluations. Report to the FAA. Available from B.G. Brown, NCAR, P.O. Box 3000, Boulder, CO 80307, 10 pp.

Drazin, P.G. and W.H. Reid, 1981: *Hydrodynamic Stability*. Cambridge, 527 pp.

Dutton, J. and H. A. Panofsky, 1970: Clear Air Turbulence: A mystery may be unfolding. *Science*, **167**, 937-944.

Ellrod, G.P. and D.I. Knapp, 1992: An objective clear-air turbulence forecasting technique: verification and operational use. *Weather and Forecasting*, **7**, 150-165.

Kane, T.L., and B.G. Brown, 2000: Confidence intervals for some verification measures – a survey of several methods. *Preprints, 15th Conference on Probability and Statistics in the Atmospheric Sciences*, Asheville, NC, 8-11 May, American Meteorological Society (Boston), 46-49.

Kronebach, G. W., 1964: An automated procedure for forecasting clear-air turbulence. *Journal of Applied Meteorology,* **3,** 119-125.

Mahoney, J.L., J.K. Henderson, and P.A. Miller, 1997: A description of the Forecast Systems Laboratory's Real-Time Verification System (RTVS). *Preprints, 7th Conference on Aviation, Range, and Aerospace Meteorology*, Long Beach, CA, American Meteorological Society (Boston), J26-J31.

Mahoney, J.L., and B.G. Brown, 2000: Forecaster Assessment of Turbulence Algorithms: A Summary of Results for the Winter 2000 Study. Report to the FAA. Available from J.L. Mahoney, FSL. 325 Broadway, Boulder, CO 80303.

Mahoney, J.L., J. Braid, B. Brown, T. Folwer, and J. Wolff, 2001a: A Forecaster Evaluation of Turbulence Algorithms. A Summary of the Winter 2001 Study. Report to the FAA. Available from J.L. Mahoney, FSL. 325 Broadway, Boulder, CO 80303, 42 pp.

Mahoney, J.L., B.G. Brown, R. Bullock, C. Fischer, J. Henderson, and B. Sigren, 2001b: Turbulence Algorithm Intercomparison: Winter 2001 Results. Report to the FAA. Available from J.L. Mahoney, FSL. 325 Broadway, Boulder, CO 80303.

Mahoney, J.L, J. K. Henderson, B.G. Brown, J.E. Hart, A. Loughe, C. Fischer, and B. Sigren, 2002: The Real-Time Verification System (RTVS) and its application to aviation weather

forecasts. *Preprints, 10th Conference on Aviation, Range, and Aerospace Meteorology*, 13-16 May, Portland, OR, American Meteorological Society (Boston), 323-326.

Marroquin, A., 1995: An integrated algorithm to forecast CAT from gravity wave breaking, upper fronts and other atmospheric deformation regions. *Preprints, 6th Conference on Aviation Weather Systems*, Dallas, TX, American Meteorological Society (Boston), 509-514.

Marroquin, A., 1998: An advanced algorithm to diagnose atmospheric turbulence using numerical model output. *Preprints, 16th Conference on Weather Analysis and Forecasting*, Phoenix, AZ, 11-16 January, American Meteorological Society (Boston).

Mason., I., 1982: A model for assessment of weather forecasts. *Australian Meteorological Magazine*, **30**, 291-303.

NWS, 1991: National Weather Service Operations Manual, D-22. National Weather Service. (Available at Website http://www.nws.noaa.gov).

Sharman, R, C. Tebaldi, and B. Brown, 1999: An integrated approach to clear-air turbulence forecasting. *Preprints, 8th Conference on Aviation, Range, and Aerospace Meteorology*, Dallas, TX, 10-15 January, American Meteorological Society (Boston), 68-71.

Sharman, R, B. Brown, and S. Dettling, 2000: Preliminary results of the NCAR Integrated Turbulence Forecasting Algorithm (ITFA) to forecast CAT. *Preprints, 9th Conference on Aviation, Range, and Aerospace Meteorology*, Orlando, FL, 11-15 Sept., American Meteorological Society (Boston), 460-465.

Sharman, R., C. Tebaldi, J. Wolff, and G. Wiener, 2002a: Results from the NCAR Integrated Turbulence Forecasting Algorithm (ITFA) for predicting upper-level clear-air turbulence. *Preprints, 10th Conference on Aviation, Range, and Aerospace Meteorology*, Portland, OR, 13-16 May, American Meteorological Society (Boston), 351-354.

Sharman, R., J. Wolff, G. Wiener, and C. Tebaldi, 2002b: Technical Description Document for the Integrated Turbulence Forecasting Algorithm (ITFA). Report, submitted to the Federal Aviation Administration Aviation Weather Research Program (FAA/AWRP), available from R. Sharman (sharman@rap.ucar.edu).

Stull, R.B., 1988: *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, p 666.

Tebaldi, C., D. Nychka, B.G. Brown, and R. Sharman, 2002: Flexible discriminant techniques for forecasting clear-air turbulence. *Environmetrics*, in press.

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

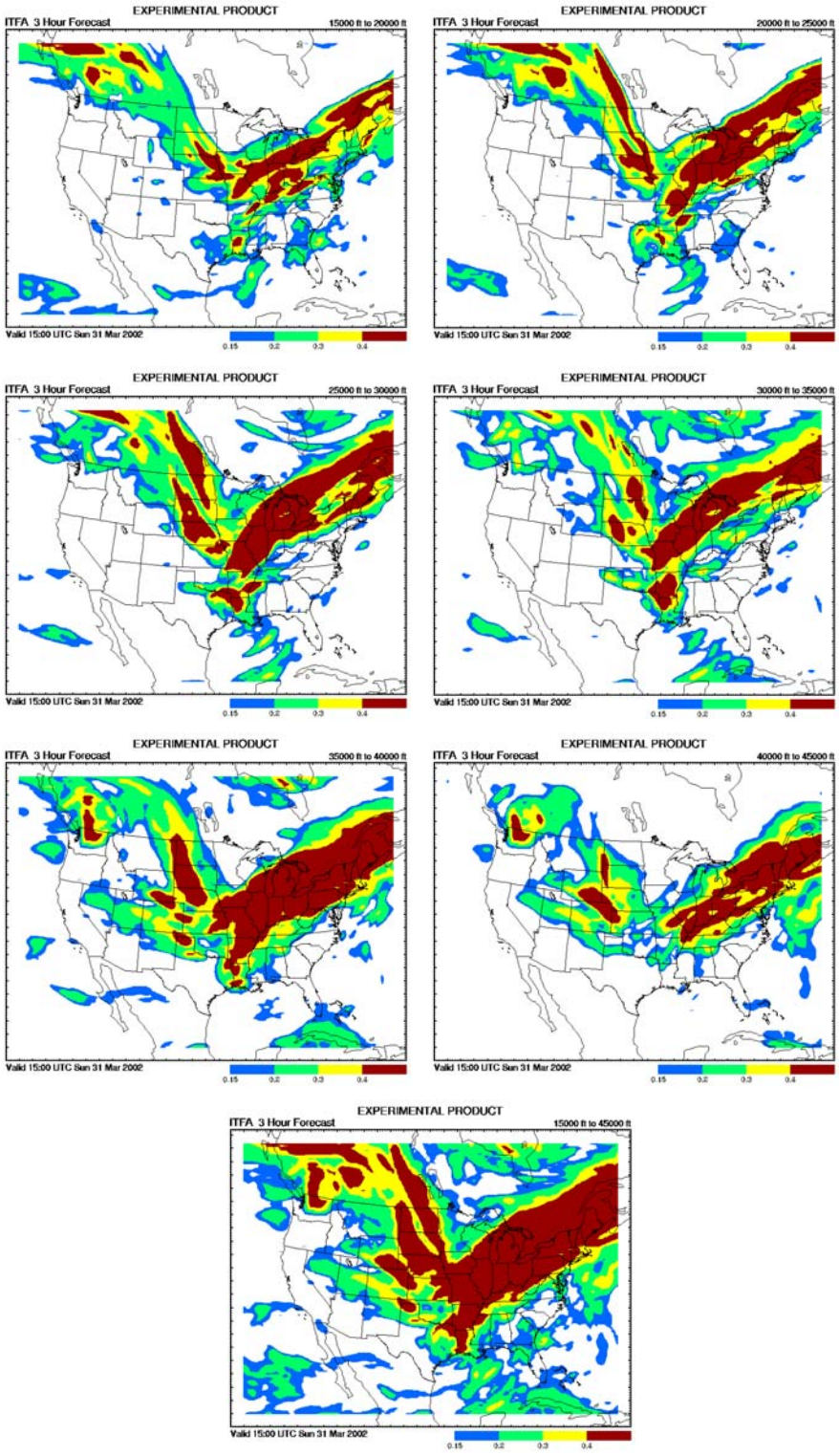

Figure 1. Example of output from ITFA, for 3-h forecast issued at 1200 UTC on 31 March 2002. Forecasts for individual layers are shown, as well as composite for 15-45,000 ft.

Figure 2. RUC-2 domain. Tics on the edges of the frame identify the model grid lines; dark outline around continental U.S. denotes the total domain of the AIRMETs.

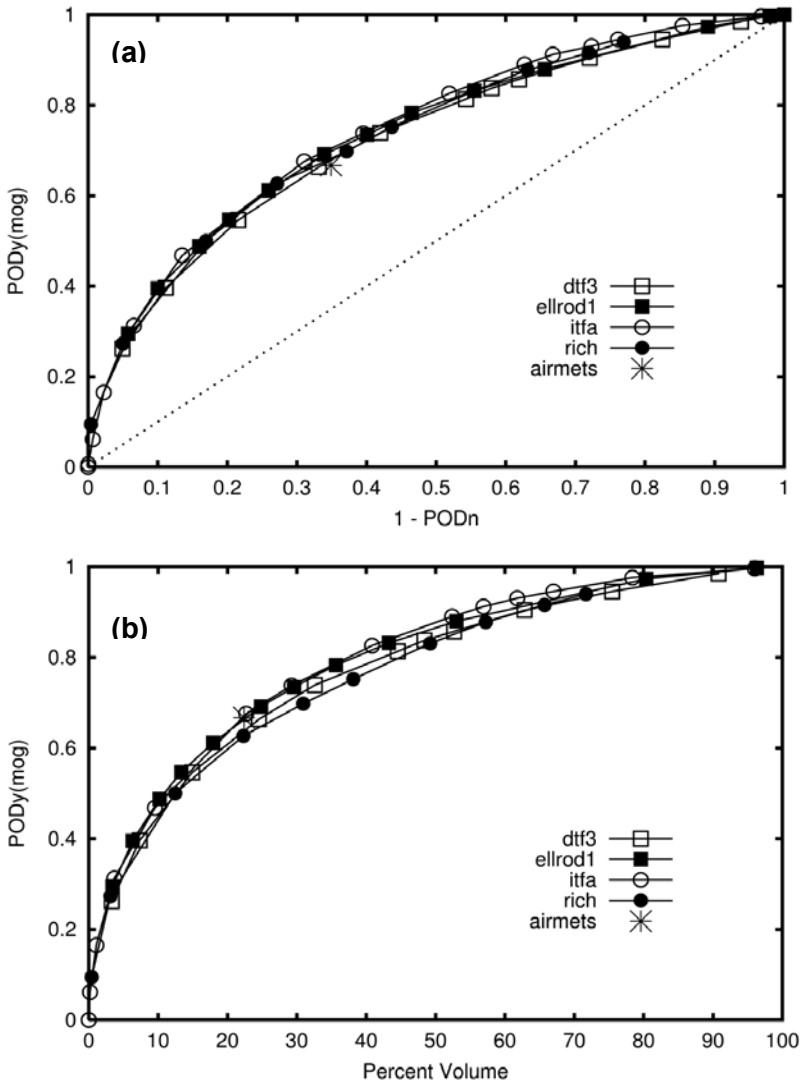

Figure 3. Overall verification statistics for 3-h forecasts, as verified in post-analysis, showing relationship between PODy(MOG) and (a) 1-PODn and (b) % Volume.

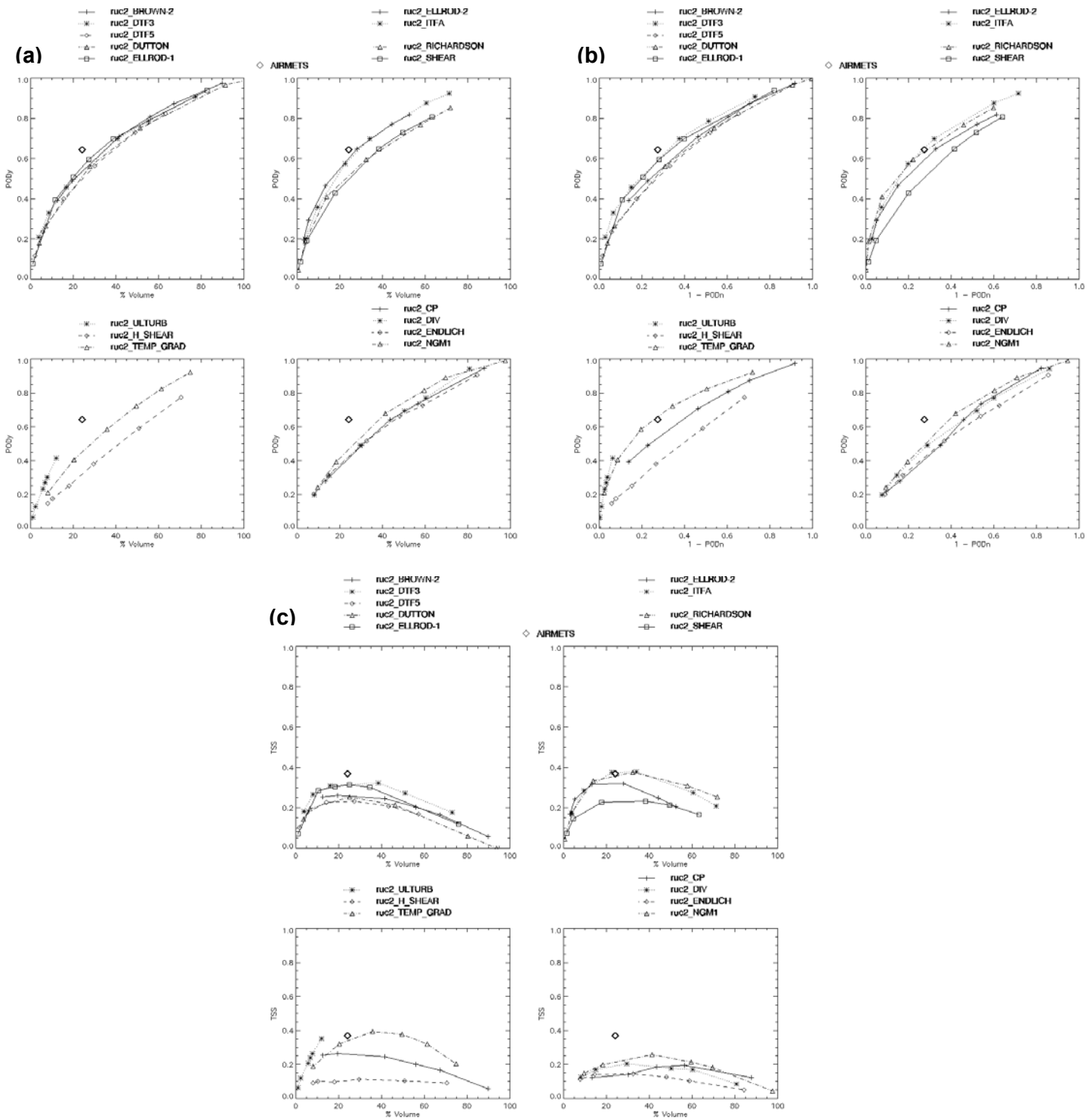

Figure 4. RTVS results for 3-h forecasts issued at 1500 UTC: (a) PODy vs. % Volume; (b)PODy vs. 1-PODn; and (c) TSS vs. % Volume.

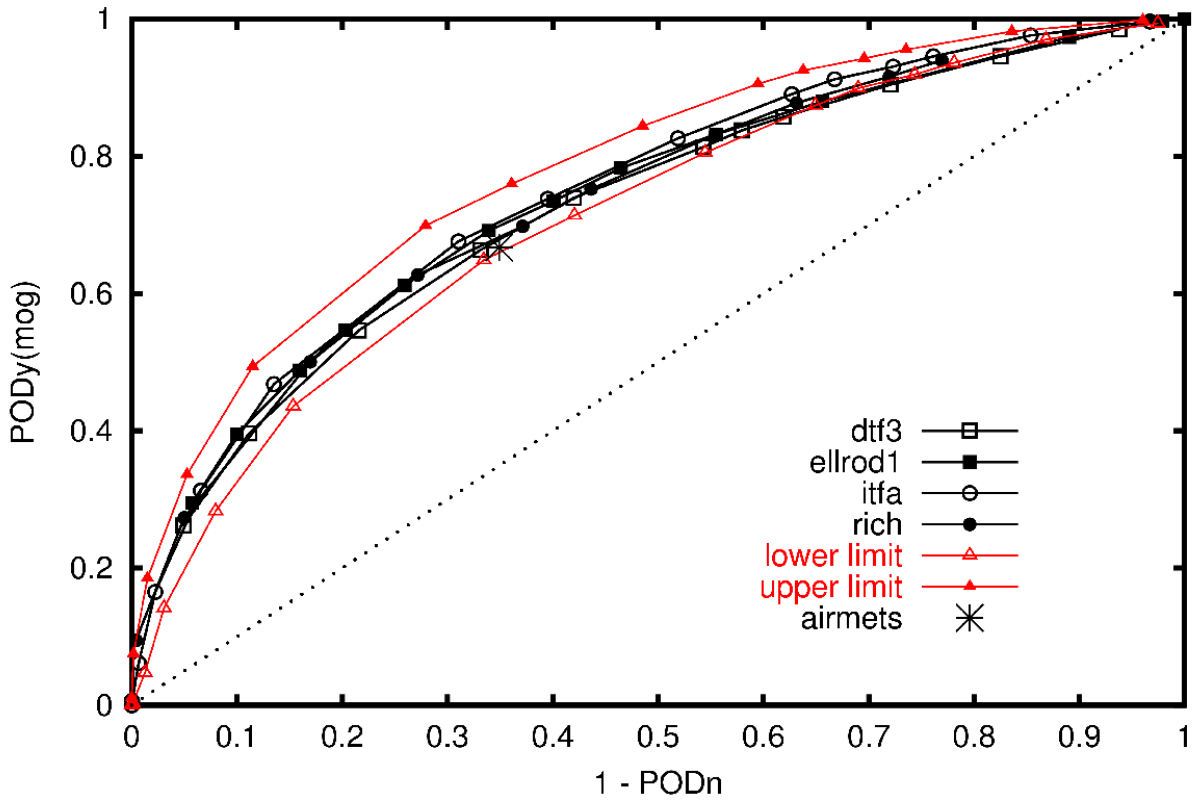

Figure 5. 95% confidence interval for ITFA PODy, along with curves for other algorithms.

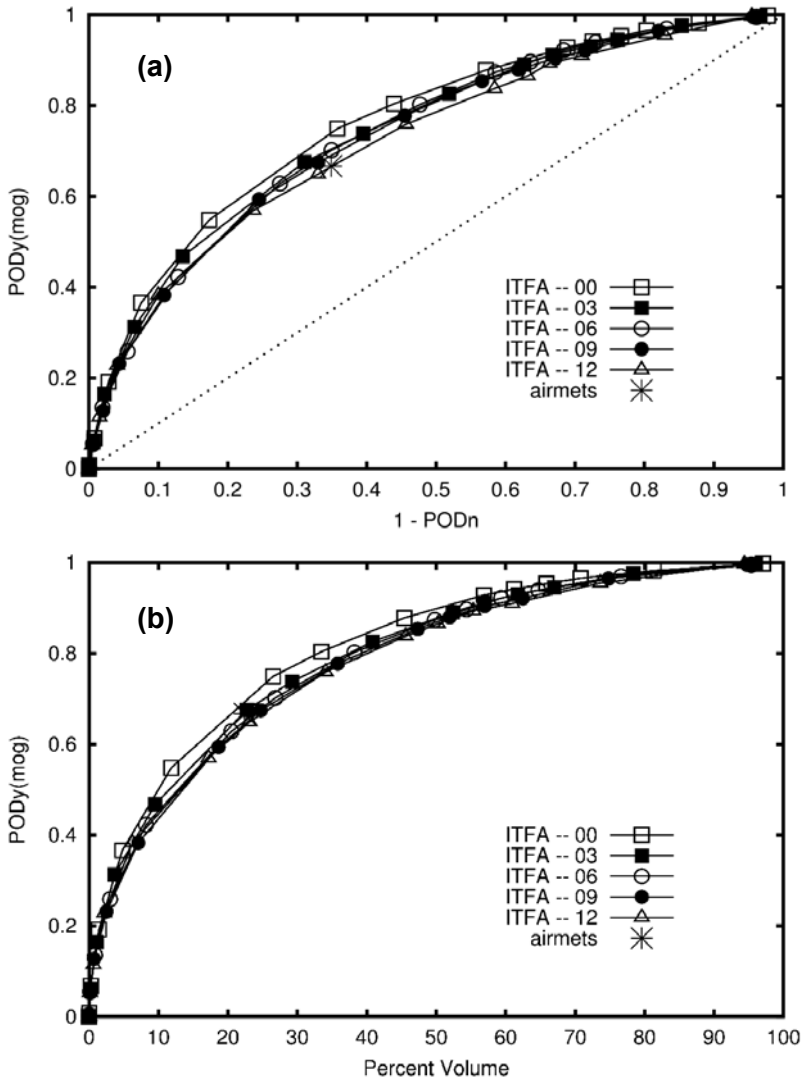

Figure 6. Variations in ITFA verification statistics with lead time for forecasts above 20,000 ft: (a) ROC diagram and (b) PODy vs. % Volume.

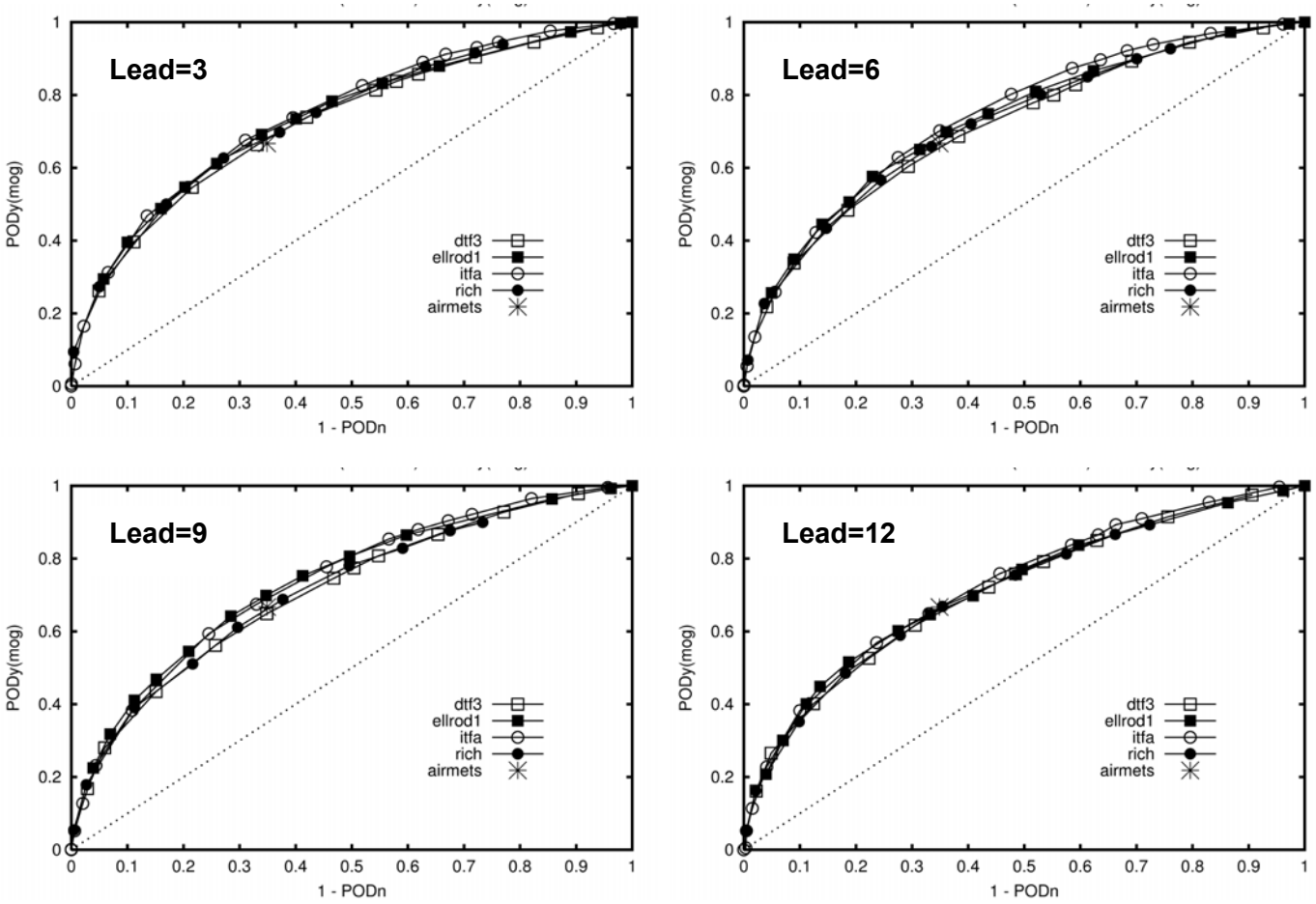

Figure 7. ROC diagrams for individual algorithms by lead time.

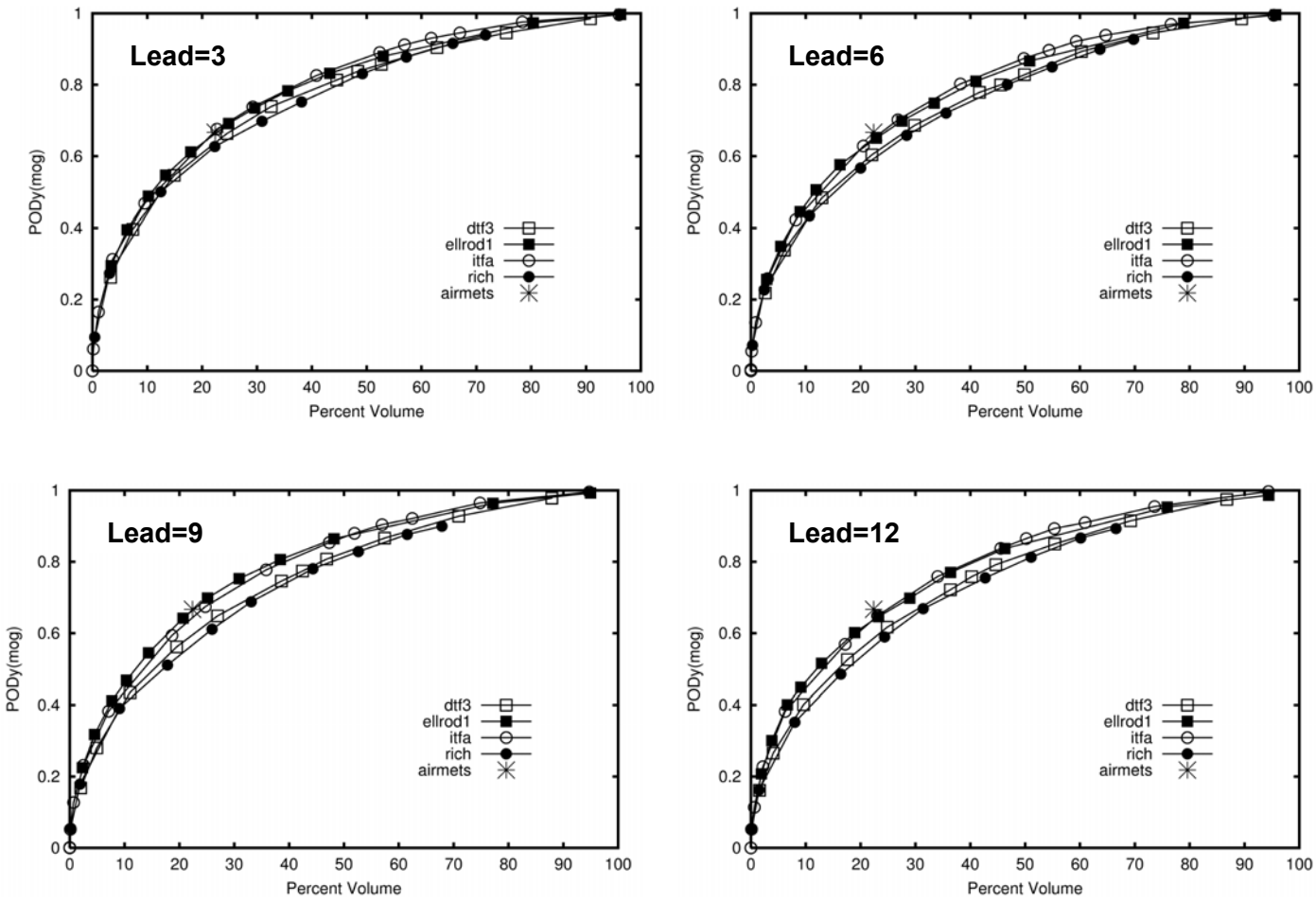

Figure 8. PODy vs % Volume plots for all of the algorithms by lead time, for altitudes of 20,000 ft and above.

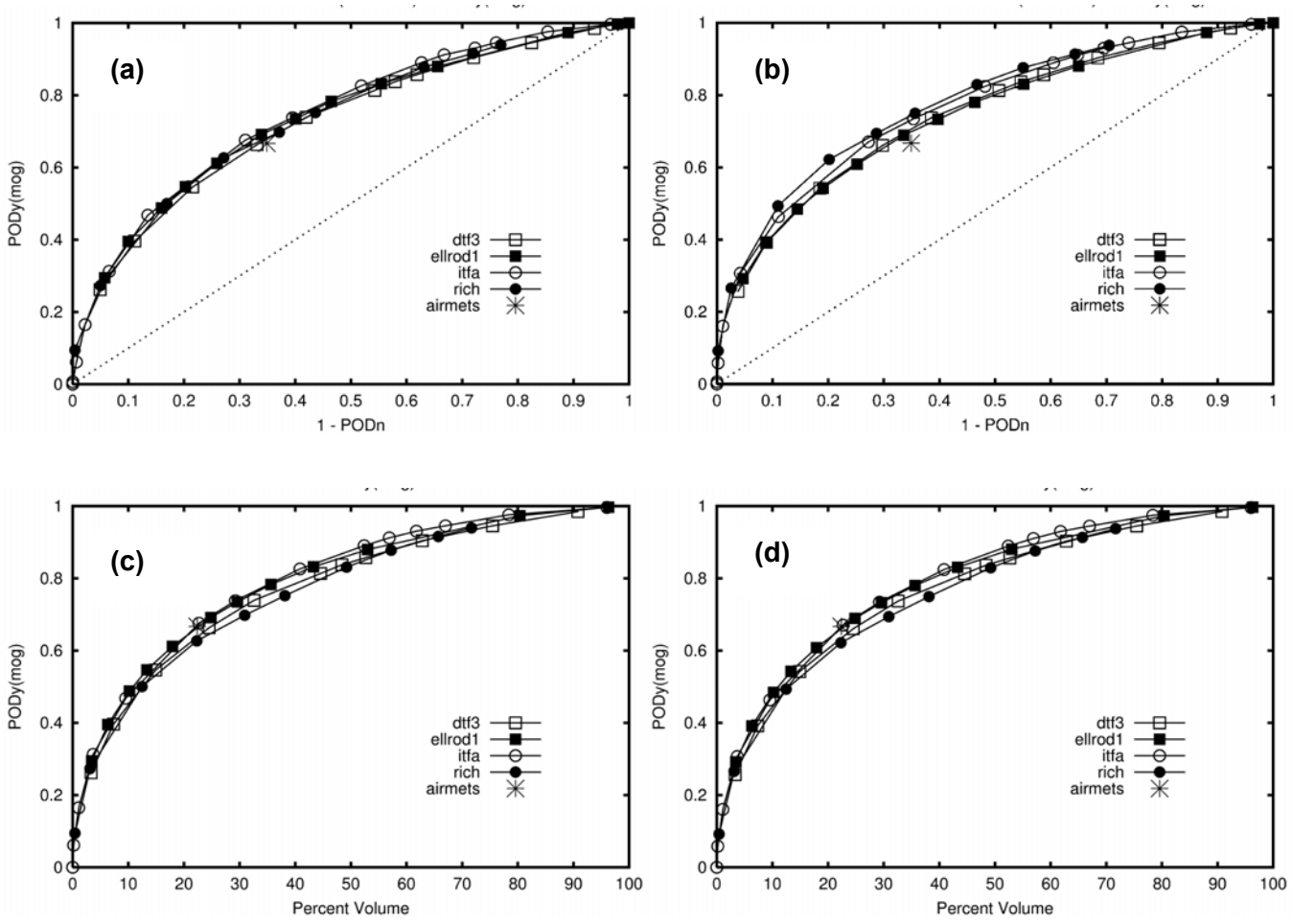

Figure 9. Variation in ROC and % Volume curves with type of PIREP: (a) and (c) regular PIREPs; (b) and (d) with supplemental UAL and NWA PIREPs

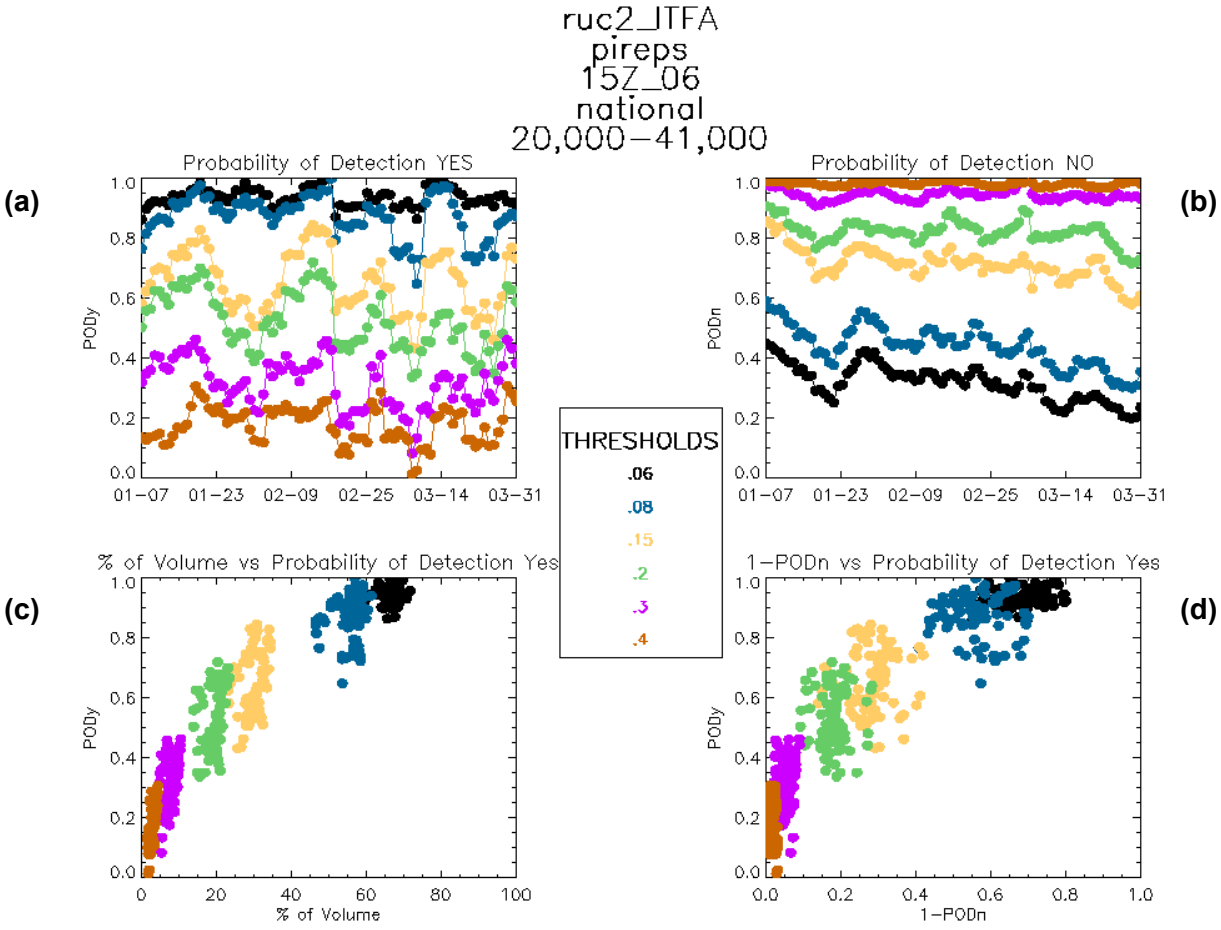

Figure 10. Variations in ITFA verification statistics for 6-h ITFA forecasts issued at 1500 UTC: (a) time series of PODy; (b) time series of PODn; (c) PODy vs. % Volume; and (d) PODy vs. 1-PODn.

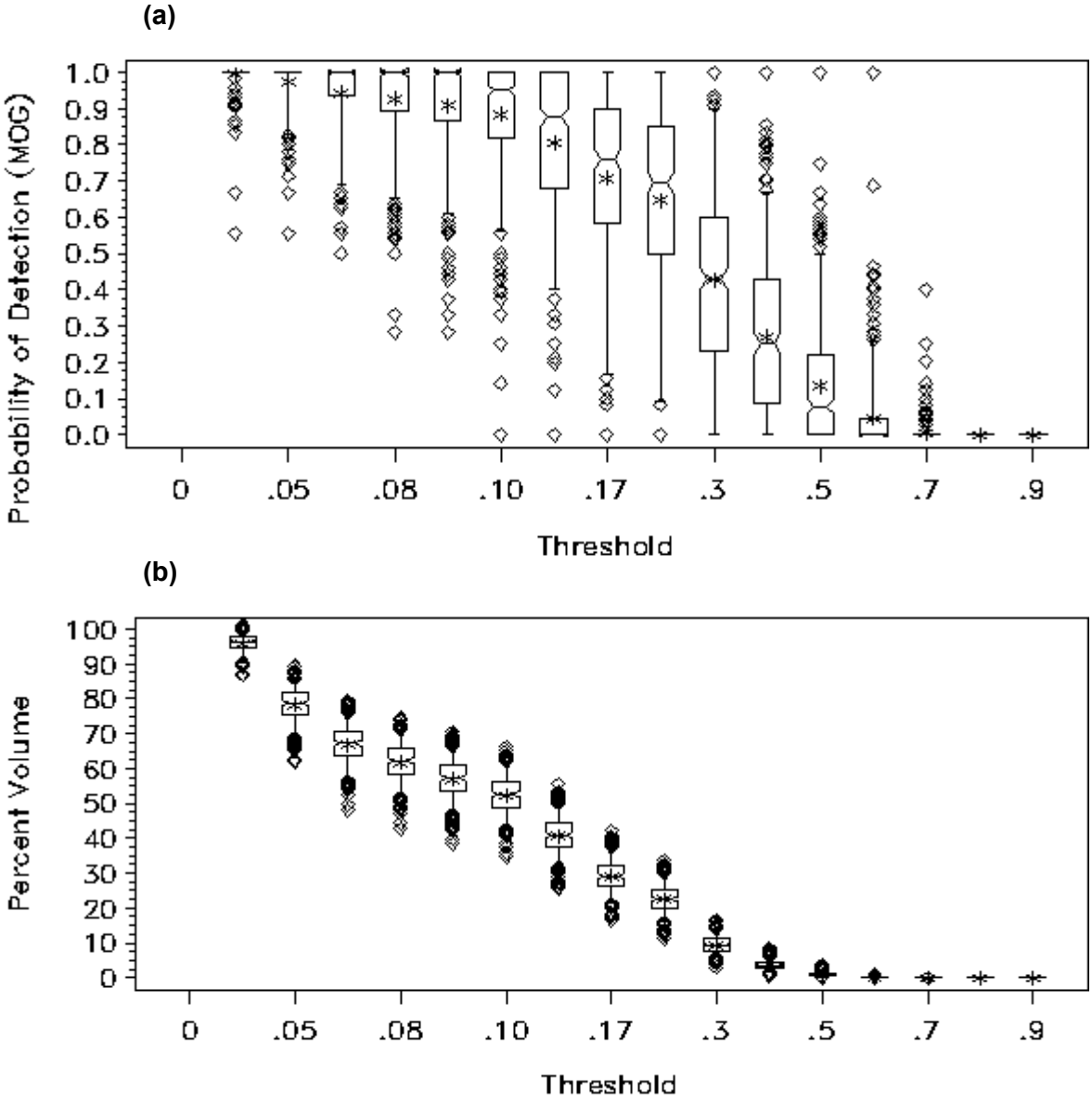

Figure 11. Box plots showing distributions of (a) PODy and (b) % Volume for 3-h ITFA forecasts, as a function of ITFA threshold. Line inside each box is the median (0.50th quantile, asterisk is the mean value; top and bottom of the box are the 0.75th and 0.25th quantiles (upper and lower quartiles) of the distribution; upper and lower "whiskers" represent the 0.95th and 0.05th quantile values. Points above and below the whiskers are the extreme values.

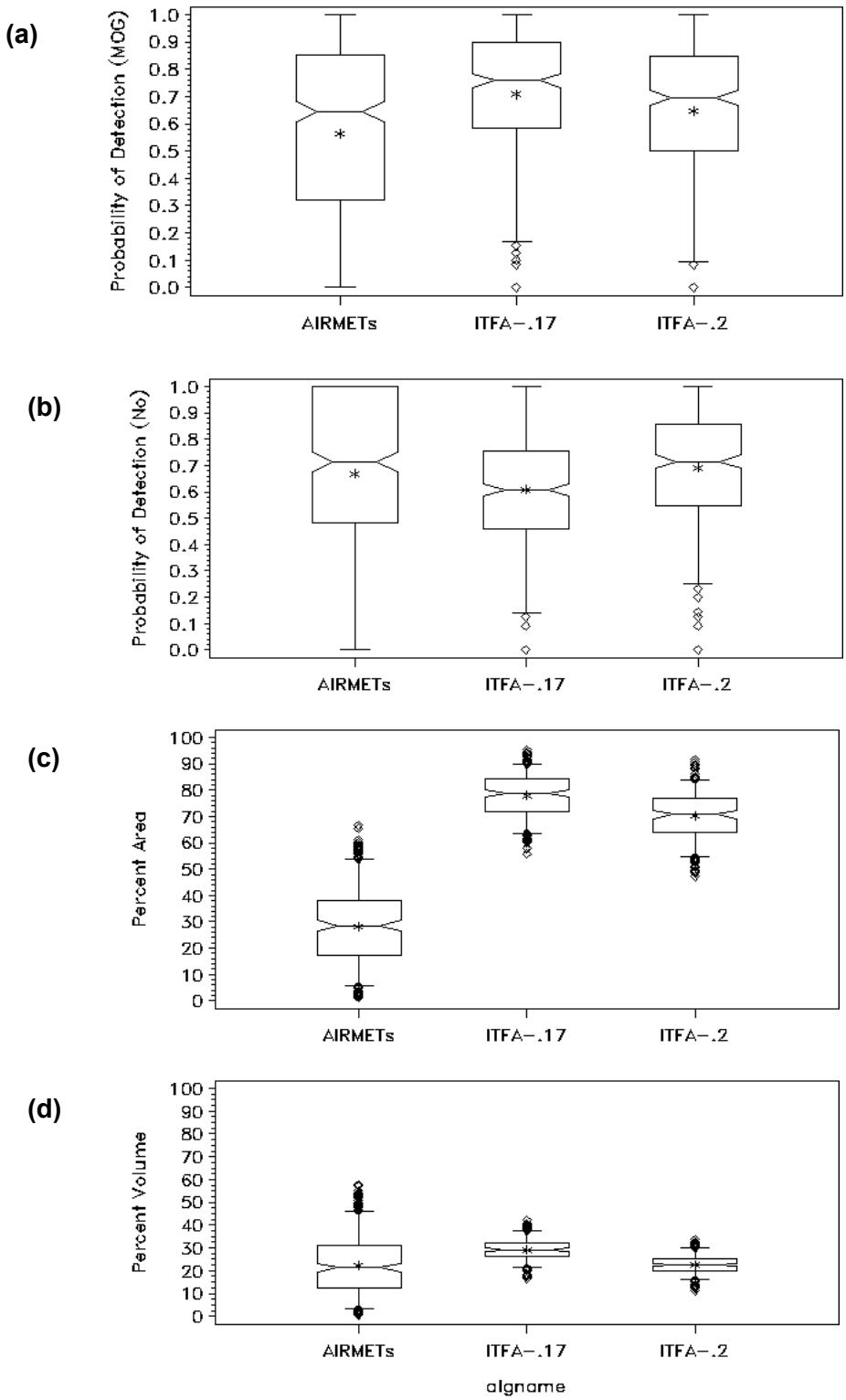

Figure 12. Box plots showing distributions of (a) PODy, (b) PODn, (c) % Area and (d) % Volume for 3-h ITFA forecasts (with two different thresholds) and AIRMETs. Box plots defined as in Fig. 11.

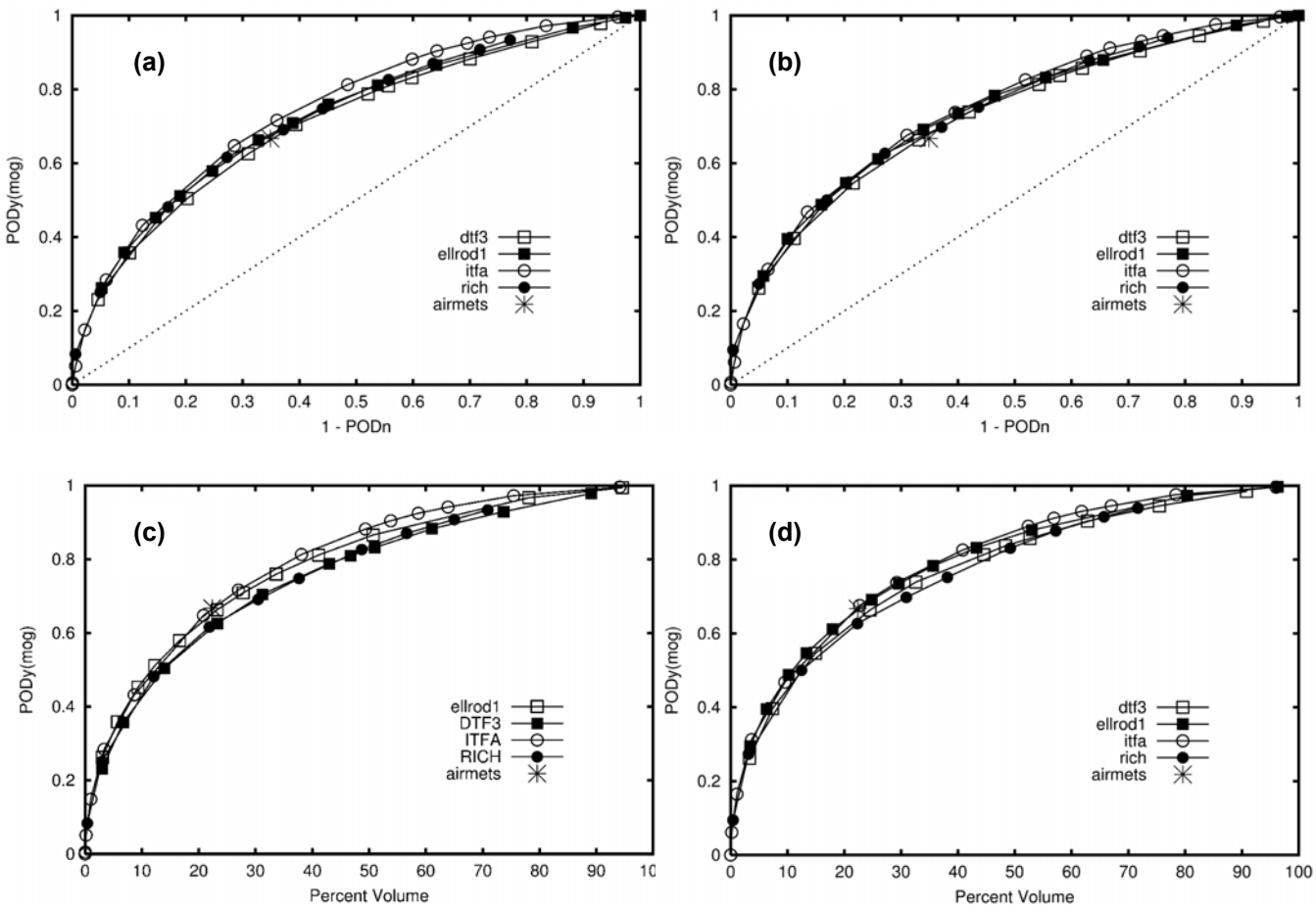

Figure 13. Variation in ROC and % Volume curves with altitude range: (a) and (c) above 15,000 ft; (b) and (d) above 20,000 ft.

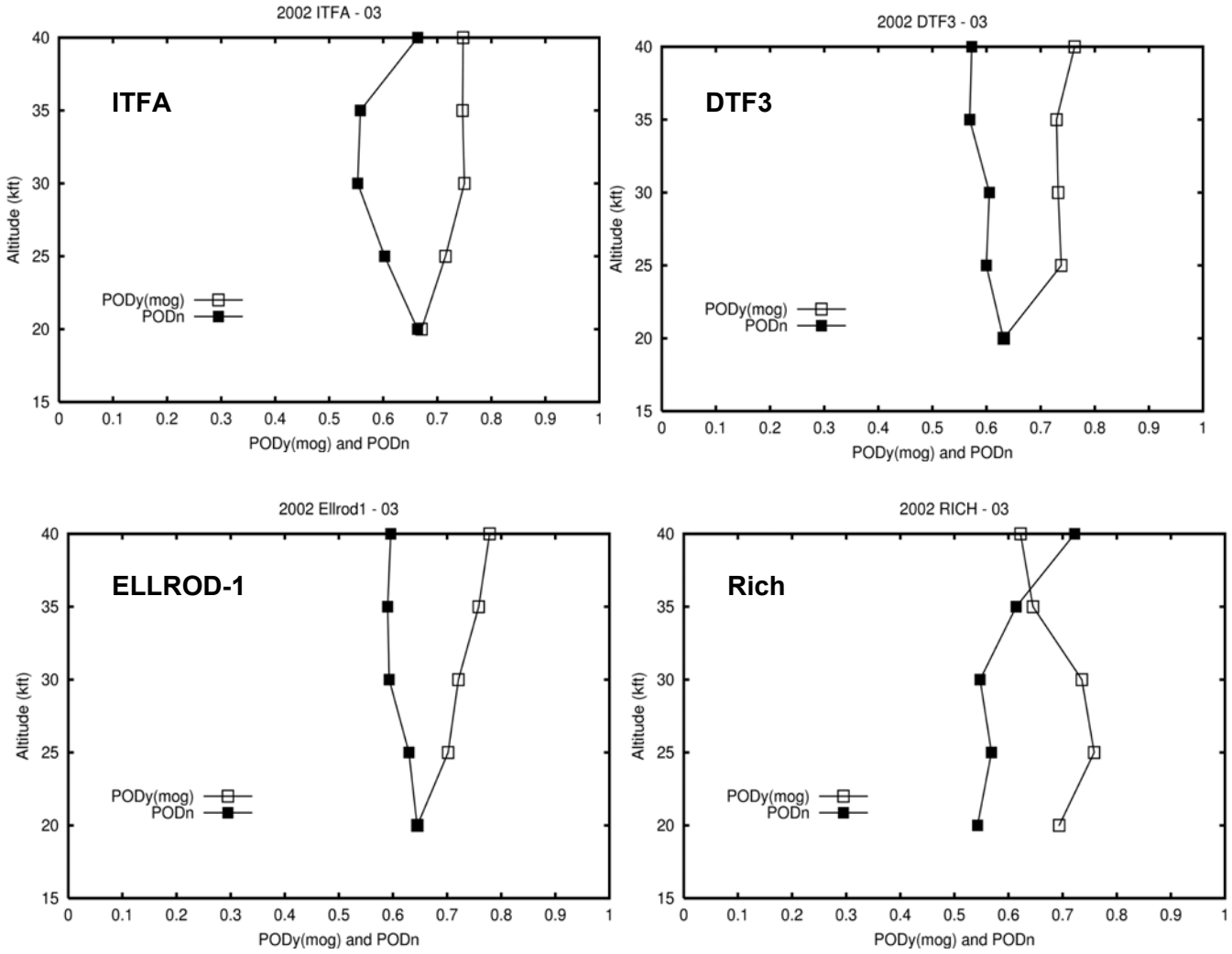

Figure 14. Variations in PODy and PODn with altitude, for ITFA and other algorithms, for 3-h forecasts.

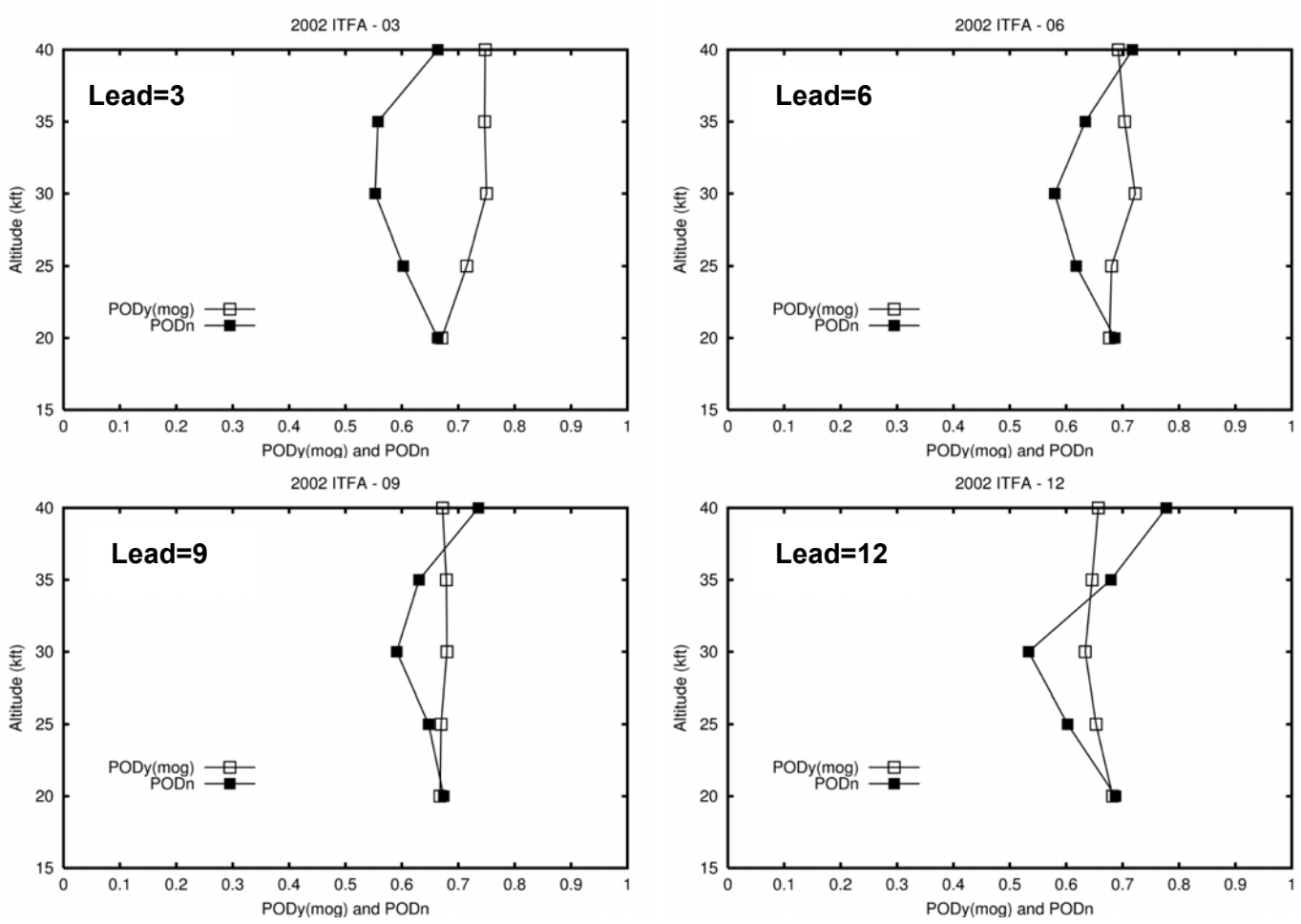

Figure 15. Variations in ITFA verification statistics (PODy and PODn) with altitude, by lead time.

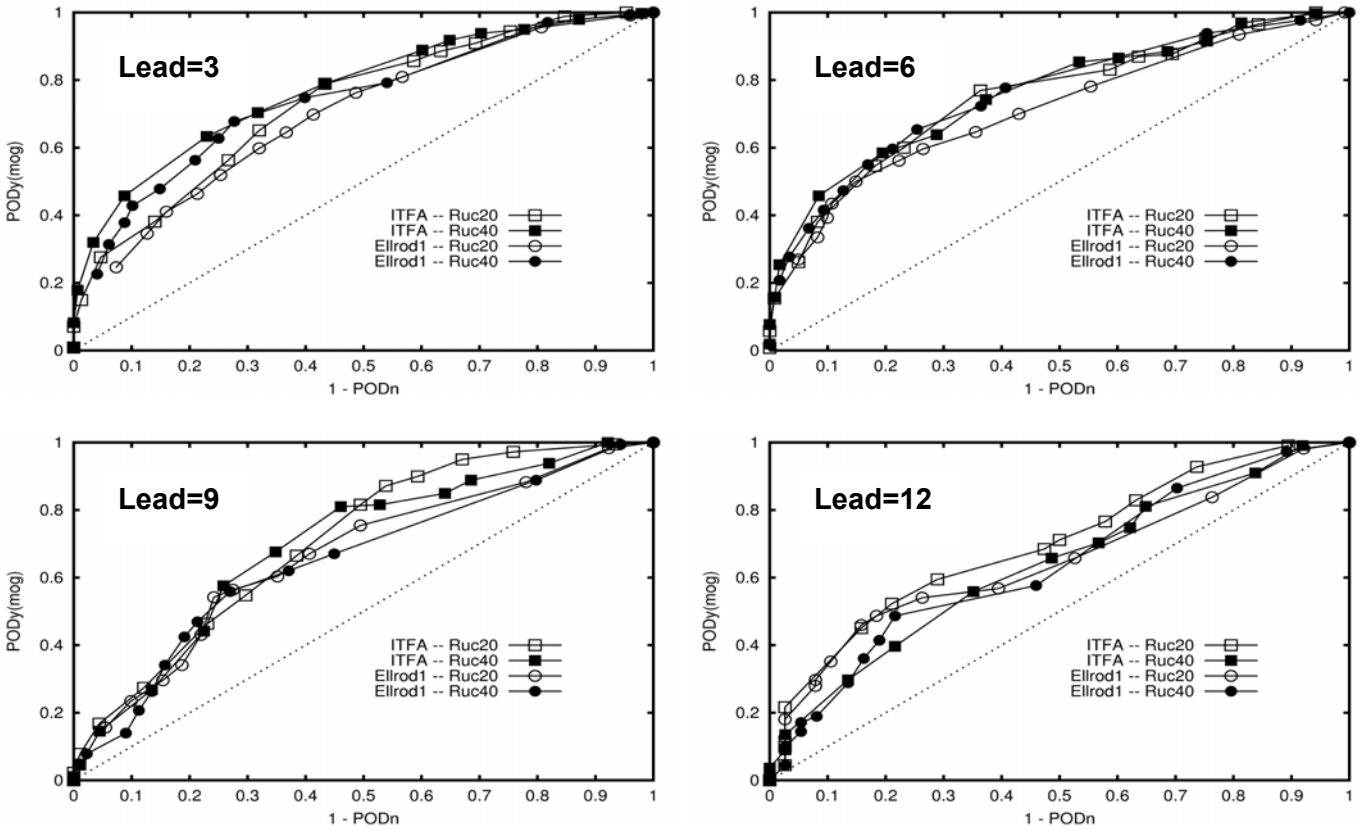

Figure 16. ROC curves for comparison of ITFA forecasts on the RUC-20 vs. the RUC-40, by lead time.

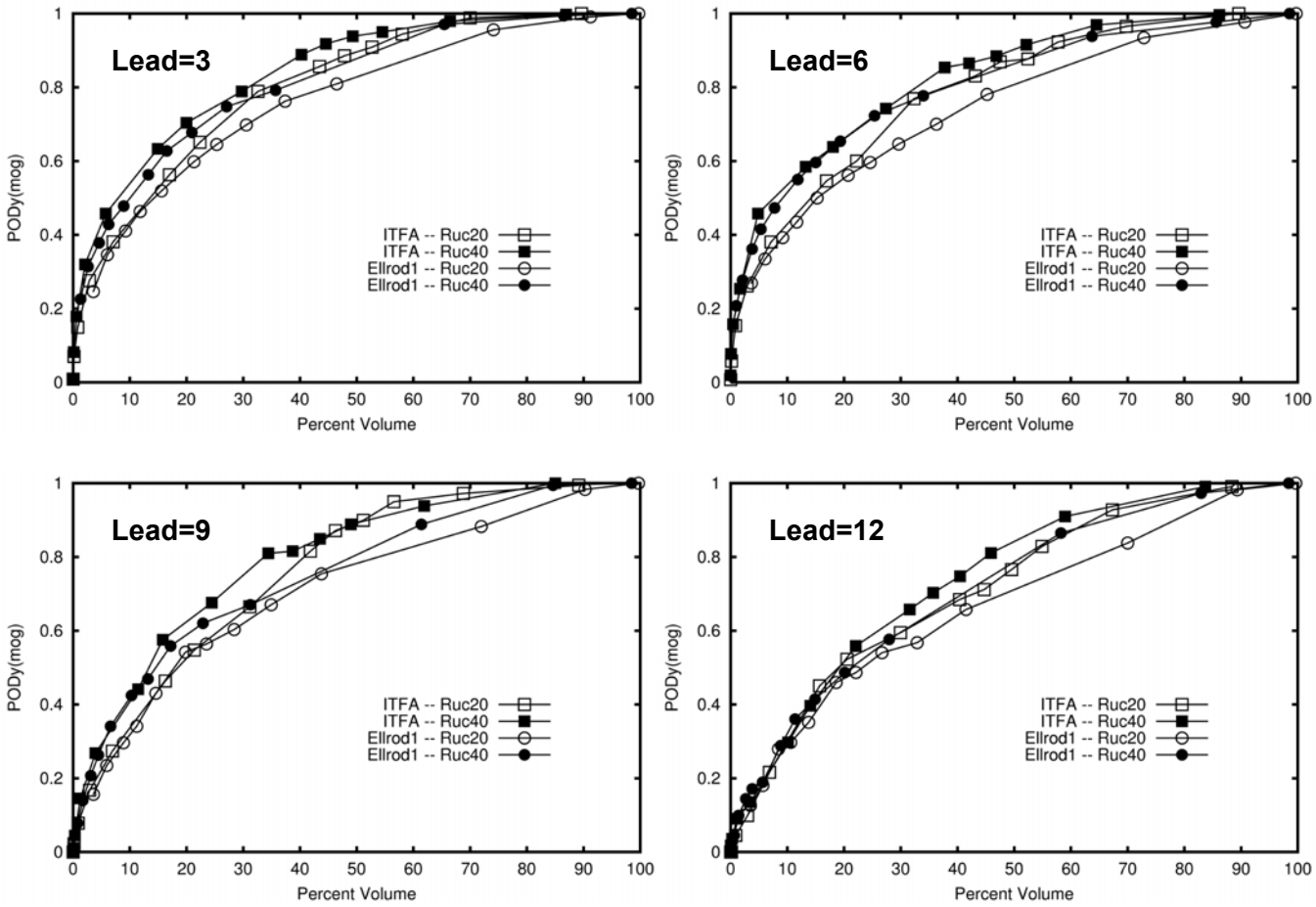

Figure 17. PODy vs. % Volume curves for comparison of ITFA forecasts on the RUC-20 vs. the RUC-40, by lead time.

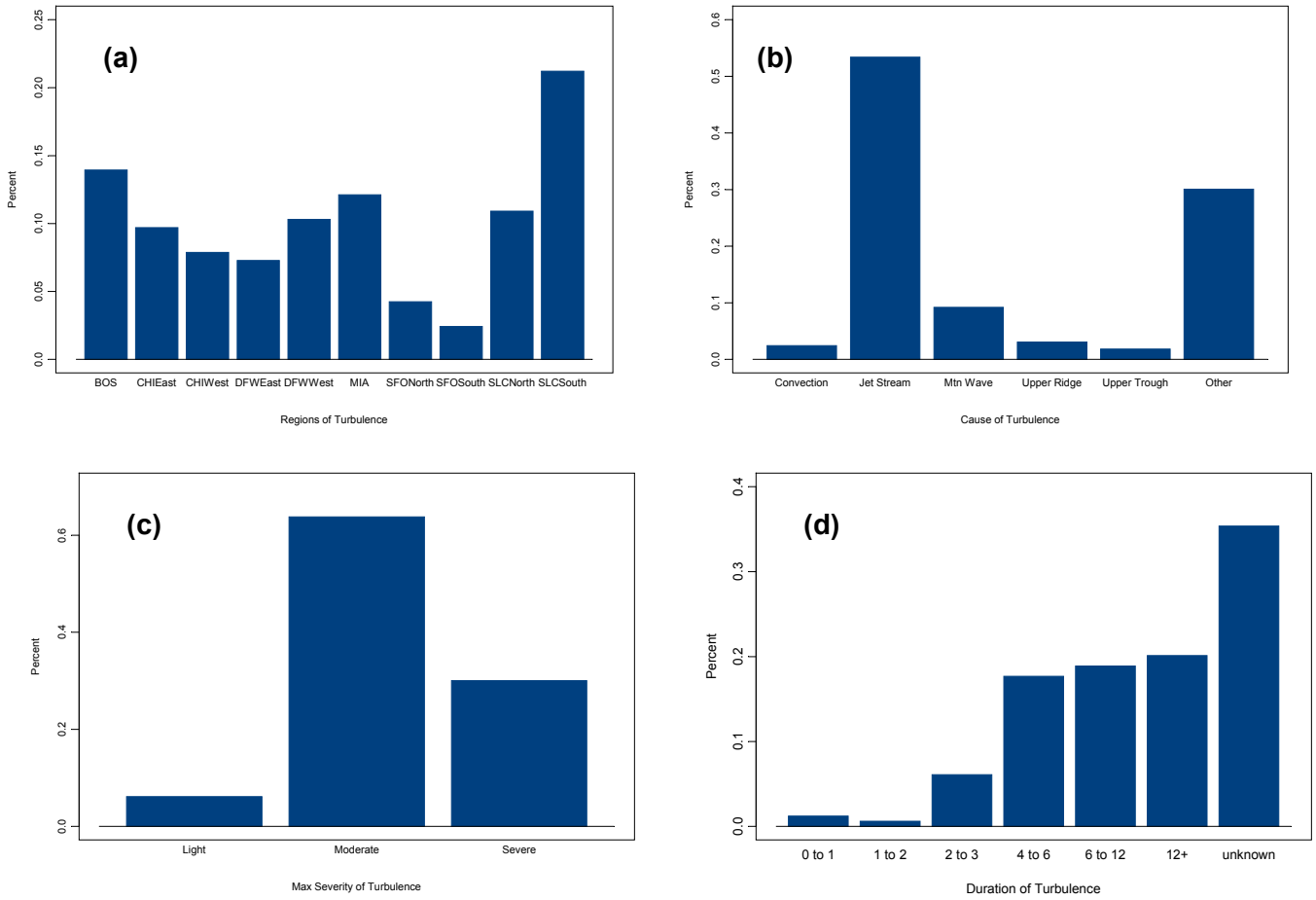

Figure 18. Characteristics of turbulence evants considered by AWC forecasters during winter 2002 subjective evaluation of ITFA: (a) region; (b) cause; (c) severity; and (d) duration.

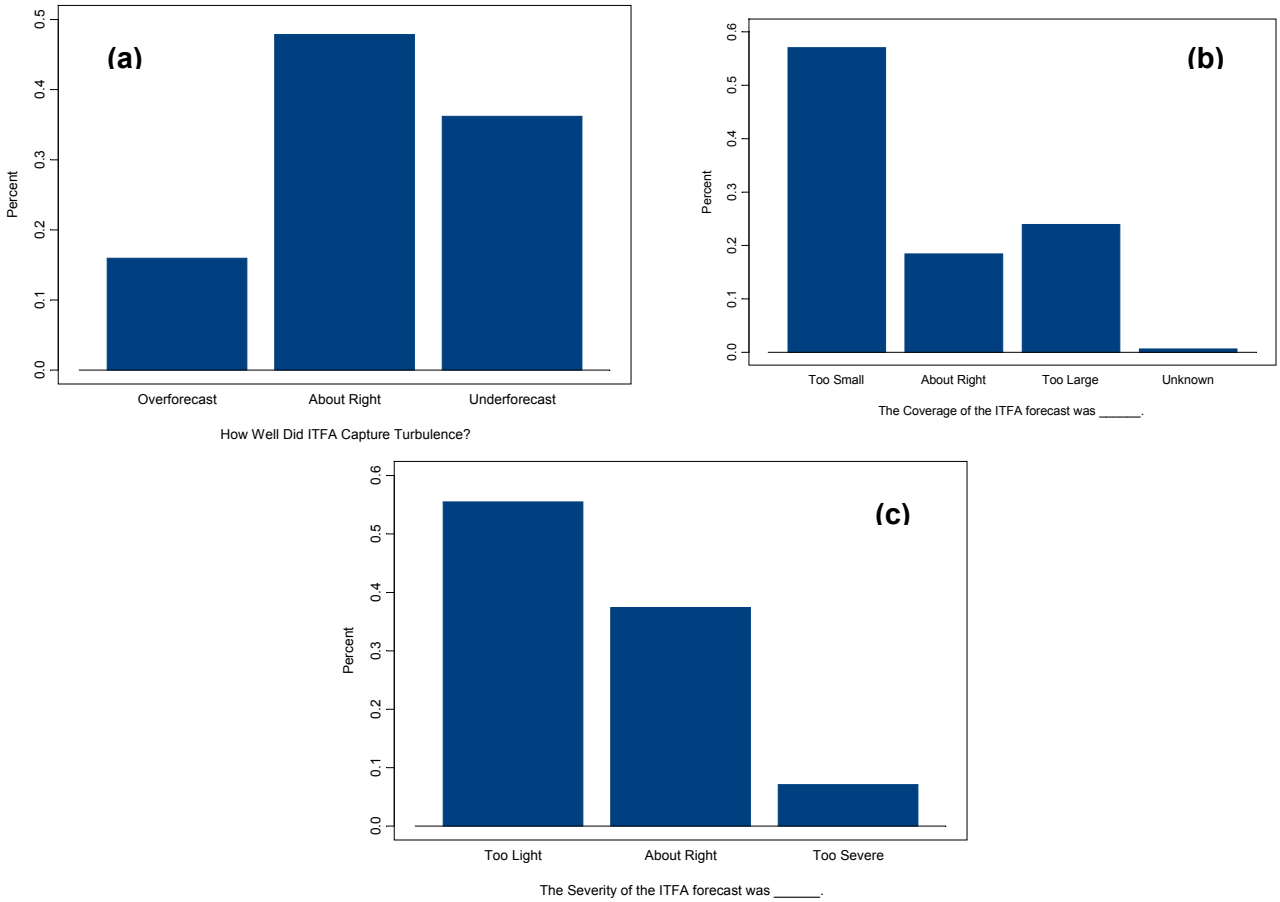

Figure 19. Responses of AWC forecasters to questions about ITFA performance: (a) "How well did ITFA capture the turbulence?"; (b) "The coverage of the ITFA forecast was ____"; and (c) "The severity of the ITFA forecast was ____".

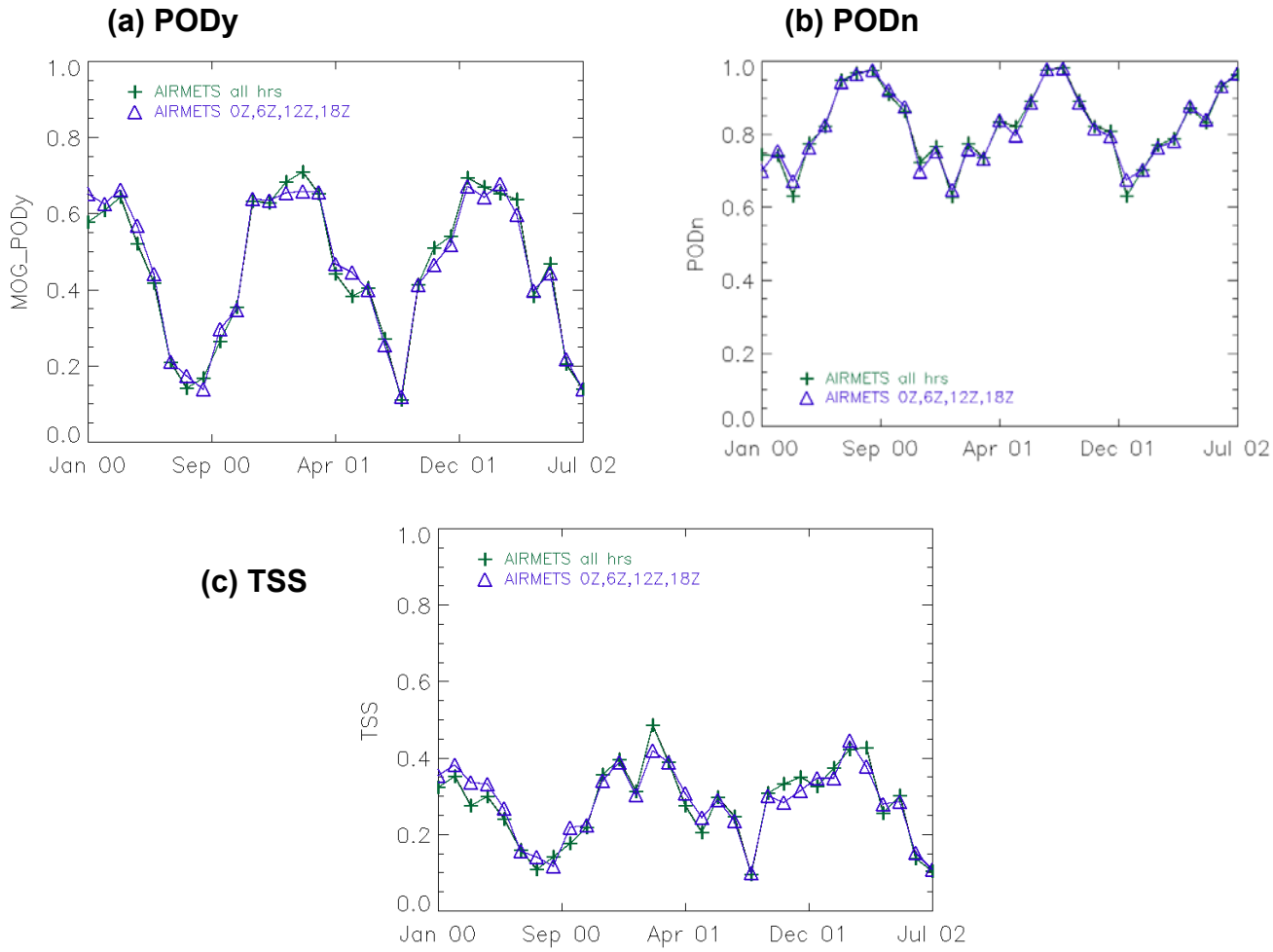

Figure 20. Time series of AIRMET verification statistics by month, for all hours combined and subset of hours (0000, 0600, 1200, and 1800 UTC): (a) PODy; (b) PODn; and (c) TSS.

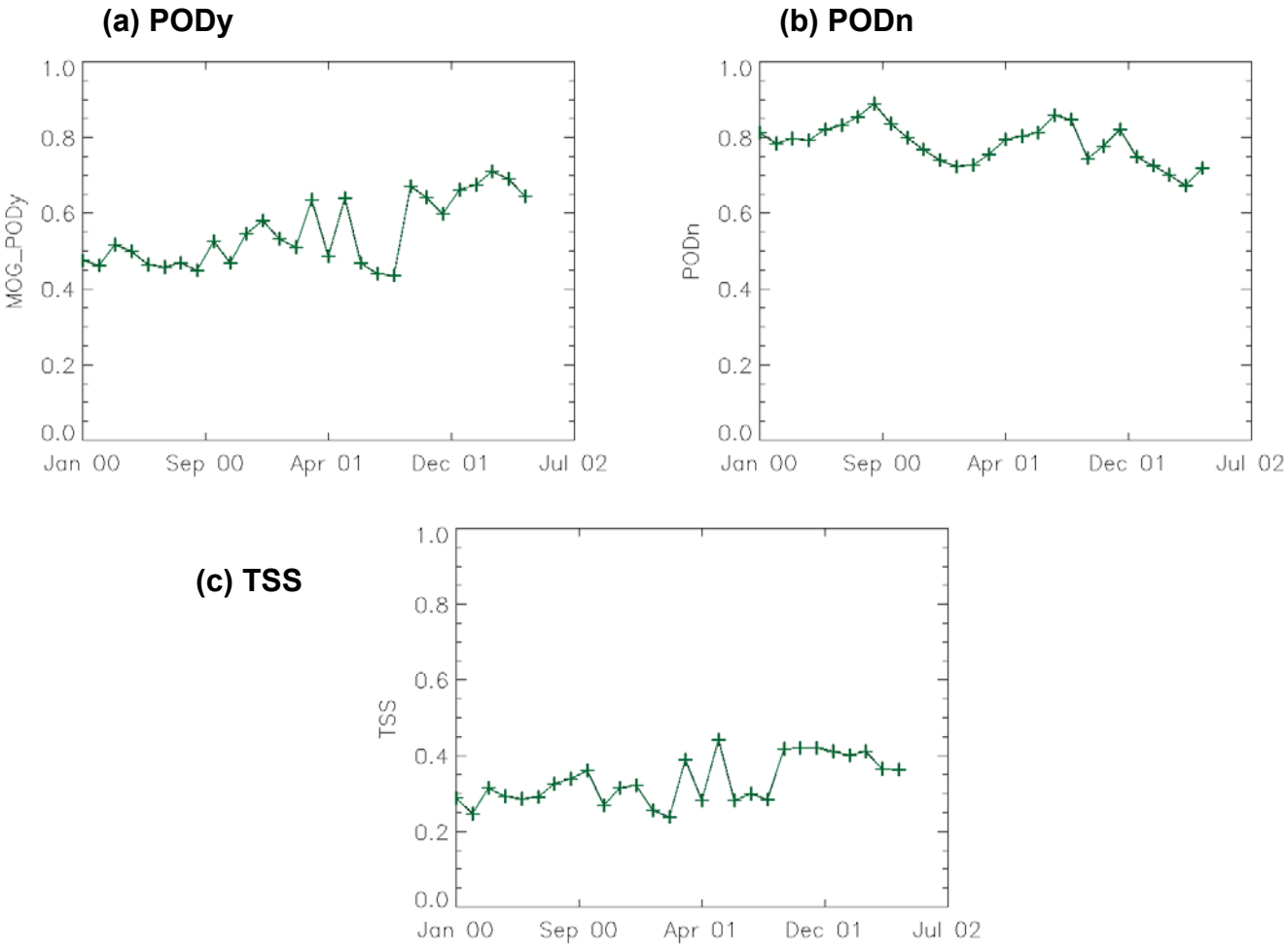

Figure 21. Time series of ITFA verification statistics by month, for ITFA threshold of 0.15, for alll lead times combined: (a) PODy; (b) PODn; and (c) TSS.