

An Integrated Approach to Mid- and Upper-Level Turbulence Forecasting

R. SHARMAN, C. TEBALDI, G. WIENER, AND J. WOLFF

Research Applications Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 6 June 2005, in final form 1 December 2005)

ABSTRACT

An automated procedure for forecasting mid- and upper-level turbulence that affects aircraft is described. This procedure, termed the Graphical Turbulence Guidance system, uses output from numerical weather prediction model forecasts to derive many turbulence diagnostics that are combined as a weighted sum with the relative weights computed to give best agreement with the most recent available turbulence observations (i.e., pilot reports of turbulence or PIREPs). This procedure minimizes forecast errors due to uncertainties in individual turbulence diagnostics and their thresholds. Thorough statistical verification studies have been performed that focused on the probabilities of correct detections of yes and no PIREPs by the forecast algorithm. Using these statistics as a guide, the authors have been able to intercompare individual diagnostic performance, and test various diagnostic threshold and weighting strategies. The overall performance of the turbulence forecast and the effect of these strategies on performance are described.

1. Introduction

Commercial, air taxi, and general aviation (GA) aircraft encounters with turbulence continue to be a source of occupant injuries, and in the case of GA, fatalities and loss of aircraft. Although the number of fatalities related to commercial airline turbulence encounters have been very low (only three in the last 10 yr), turbulence encounters do account for a significant percentage (about 65%) of all weather-related commercial aircraft incidents. The average number of air carrier turbulence-related injuries is about 45 per year according to the National Transportation Safety Board (NTSB), but these are of course only those cases that were reported to the NTSB. The actual number is probably much higher: one major carrier reported almost 400 injury-causing turbulence encounters over a 3-yr period; another estimated about 200 turbulence-related customer injury claims per year. Over the last 12 yr, the average number of moderate-or-greater and severe-or-greater encounters of turbulence actually reported and

recorded amounts to over 63 000 and 5000 per year, respectively. Costs to the airlines are difficult to establish, but a vice president for one major air carrier, in a presentation delivered at the National Aeronautics and Space Administration–Federal Aviation Administration (NASA–FAA)-sponsored Aircraft Turbulence Accident Prevention First Users’ and Technologists’ Workshop in Hampton, Virginia, in 1998, estimated that it pays out “tens of millions per year” for customer injuries, and loses about 7000 days in employee injury-related disabilities. The vast majority of air carrier turbulence incidents occur above 10 000 ft, where passengers and flight attendants are more likely to be unbuckled.

A large number of these turbulence encounters might be avoided if better turbulence forecast products were available to air traffic controllers, airline flight dispatchers, and flight crews. In fact, previous studies (e.g., Fahey 1993) have shown that for commercial air carriers, strategic planning to avoid turbulence encounters can lead to a reduction in cabin injuries and costs. However, current forecasting methods have not generally provided acceptably high detection rates and at the same time acceptably low false alarm rates to achieve significant reductions. The term “acceptable” does not have a universal quantitative definition, but the Turbulence Joint Safety Implementation Team, a team of representatives from the FAA, NASA, various federal laboratories, and end users, recommended probabilities of moderate or greater (MOG) turbulence detection

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Dr. Robert Sharman, Research Applications Laboratory, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: sharman@ucar.edu

should be >0.8 and probabilities of null detection should be >0.85 for turbulence forecasts to be most useful. These goals are currently not achievable by either automated or experienced human forecasters.

The turbulence forecasting difficulty is due in large part to the fact that, from the meteorological perspective, turbulence is a “microscale” phenomenon. In the atmosphere, turbulent “eddies” are contained in a spectrum of sizes, from 100s of kilometers down to centimeters. But aircraft bumpiness is most pronounced when the size of the turbulent eddies encountered is about the size of the aircraft; for commercial aircraft this would correspond to eddy dimensions of ~ 100 m. It is impossible to directly and routinely forecast atmospheric motion at this scale, now or even in the foreseeable future. Fortunately, it appears that most of the energy associated with turbulent eddies on this scale cascade down from the larger scales of atmospheric motion [e.g., Dutton and Panofsky (1970) and more recently Tung and Orlando (2003) and Koshyk and Hamilton (2001)], and these larger scales may be resolved by current weather observation networks and numerical weather prediction (NWP) models. Assuming the large-scale forecasts are sufficiently accurate, the turbulence forecasting problem is then one of identifying large-scale features that are conducive to the formation of aircraft-scale eddies.

Empirically based linkages between large-scale atmospheric features (i.e., observable by routine meteorological observations and resolvable by NWP models) and aircraft-scale turbulence (i.e., forecasting “rules of thumb”) have been developed over the years by National Weather Service (NWS) and airline meteorological forecasters. The successful application of these rules, however, depends on the forecaster, and any perceived skill diminishes rapidly with forecast lead time. Because there is now a tremendous amount of meteorological data available to forecasters, more than can be digested in a reasonable length of time, automated turbulence forecasting tools could aid humans in making decisions about where to locate regions of potential turbulence that may be hazardous to aircraft.

To address the need for an automated turbulence forecasting tool, the Research Applications Laboratory at the National Center for Atmospheric Research (NCAR/RAL) and the National Oceanic and Atmospheric Administration’s (NOAA) Earth System Research Laboratory/Global Systems Division (NOAA-Research-ESRL/GSD), under sponsorship from the FAA’s Aviation Weather Research Program, have been developing and testing a completely automated turbulence forecasting system. This system was origi-

nally dubbed the Integrated Turbulence Forecasting Algorithm (ITFA; Sharman et al. 1999, 2002) and concentrated only on the prediction of clear-air turbulence (CAT) related to jet streams and fronts at upper-levels [flight pressure altitudes $> 20\,000$ ft MSL or “flight levels”¹ (FLs) > 200]. The term “integrated” was used to describe the blending of NWP model based turbulence diagnostics with available turbulence observations used to produce the forecasts. The ITFA system became operational for qualified meteorologists and dispatchers to use as a guide for making turbulence avoidance decisions in March 2003 and at that time was renamed the Graphical Turbulence Guidance (GTG) product. The two Gs emphasize the nature of the turbulence product: “graphical” because, as opposed to traditional AIRMET (Airmen’s Meteorological Information) and SIGMET (Significant Meteorological Information) polygons, the output is provided as Web-based contours of turbulence potential, and “guidance” stresses that the output should be used as a decision support tool in addition to other information that may not be available to GTG (e.g., satellite imagery). The first generation GTG, or GTG1, provides gridded CAT forecasts stratified by flight levels with graphical displays of turbulence potential provided on NOAA’s Aviation Digital Data Service (ADDS) Web site: (information online at <http://adds.aviationweather.gov/turbulence>). An example GTG1 image from the ADDS Web site is provided in Fig. 1.

This has been followed up by a second version, GTG2, which expands the capabilities of GTG1 by extending the turbulence forecasts down to FL100 and includes some diagnostics for mountain wave-related turbulence. The FL100–FL200 altitude band is especially significant for air taxis. Thus, in the new system, there are turbulence predictions at both upper ($>FL200$) and midlevels (FL100–FL200).

Both subjective and objective evaluations of the ITFA/GTG system based on comparisons with available turbulence reports (or pilot reports, PIREPs) have been an integral part of the algorithm’s development since its inception. Independent *subjective* evaluations include those from the Aviation Weather Center (AWC) for the 2000–2003 winter seasons (Kelsch et al. 2004), Delta Airlines’s Meteorology Department (winter 2001), United Airlines’s Meteorology Department [winter 2002 and 2003; Kelsch et al. (2004)], Comair Dispatch (winter 2002), and the FAA’s William J.

¹ Flight levels are actually isobaric surfaces that correspond to a particular geopotential altitude according to the *U. S. Standard Atmosphere*.

Turbulence forecast at FL300

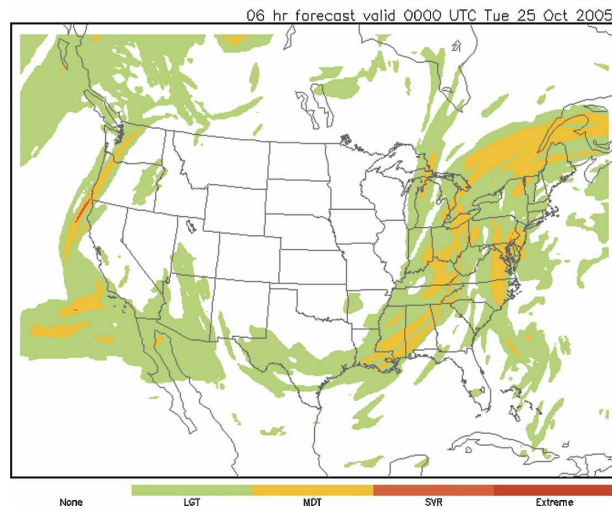


FIG. 1. Sample of GTG1 6-h forecast of CAT potential for 0000 UTC 25 Oct 2005 as provided on the operational ADDS Web site (<http://adds.aviationweather.gov/turbulence>). Color table for contours is provided at the bottom of the image. For this particular situation no turbulence is predicted over most of the western and midwestern United States, with some light to moderate turbulence regions expected on both the east and west coasts, and possibly a small region of moderate to severe turbulence off the northern California coast.

Hughes Technical Center (winter 2000) severe case studies (Passetti et al. 2000; Weinrich and Sims 2002).

Objective evaluations based on comparisons with PIREPs have been on going by both the developers and an independent verification team composed of researchers from NOAA-Research-ESRL/GSD and NCAR/RAL. Results of evaluations from previous years can be found in Brown et al. (2000). Complete objective evaluations in the form of probability of detection (POD) statistics are also available on a daily basis on NOAA-Research-ESRL/GSD's Real-Time Verification System (RTVS) (see Mahoney et al. 2002 for a description) Web site (<http://www-ad.fsl.noaa.gov/fvb/rtps>) since 1999.

This paper describes the current GTG algorithm (GTG2) and provides some statistical evaluations of its performance. The GTG methodology will be described in section 2. GTG performance derived from 1 yr (2003) of evaluations against turbulence PIREPs is presented in section 3. Development and tuning is an ongoing task, and current problem and work areas are outlined in section 4.

2. GTG procedure

The GTG process starts by automatically ingesting gridded NWP data, which are supposed to accurately represent the large-scale features of the atmosphere that may be related to aircraft-scale turbulence. In prin-

ciple, any NWP model could be used, but the National Centers for Environmental Prediction's (NCEP's) Rapid Update Cycle (RUC-2) model was chosen because of the higher effective vertical resolution provided by the isentropic vertical coordinate system at upper levels in the model (Benjamin et al. 2004). The essence of the GTG forecasting method is to integrate a combination of many separate turbulence diagnostics, with each diagnostic weighted to get the best agreement with available observations (i.e., PIREPs). This idea of using a weighted combination of diagnostics to provide turbulence forecasts is not in itself a new one. For example, Dutton (1980) evaluated the performance of 11 diagnostics compared with pilot reports of CAT over the North Atlantic and parts of Europe. He found the weighted sum of the vertical and horizontal wind shears provided the best agreement with his observations. Also Clark et al. (1975) used a set of five weighted diagnostics, where the set used depended on elevation bands and the weights were determined by the best fit to data from several XB-70 stratospheric turbulence encounters over the western United States. Similar procedures have been used by Russian investigators. For example, Leshkevich (1988) used a weighted sum of 12 diagnostics, and Buldovskii et al. (1976) used a weighted combination of horizontal temperature gradient and vertical wind shear to predict CAT, again with the weights determined by best agreement to available observations. However, all of these studies were based on a limited set of observations and the weights determined by the best fit to this limited set. These weights, once established, are static; that is, they never change. The GTG procedure also obtains weights for a set of diagnostics based on the best fit to observations, when a sufficient number of PIREPs are available in real time, and are determined dynamically and updated with every RUC model update. Alternatively, a set of climatologically derived static weights can be used when the number of observations is insufficient for robust assessment of the dynamic weights. In particular, PIREPs undergo a strong diurnal period with considerably fewer at night, roughly 0200–1300 UTC (see Fig. 2), making it difficult to use the dynamic weighting method during those times. In the following, the GTG version that makes use of the dynamic weighting strategy will be referred to as GTGD, and the version that uses the climatological weights will be referred to as GTGC. The entire GTG process involves a six-step procedure described in the following sections.

a. Step 1

In step 1 a set of n turbulence indices or diagnostics D_n (e.g., a local Richardson number) is computed from

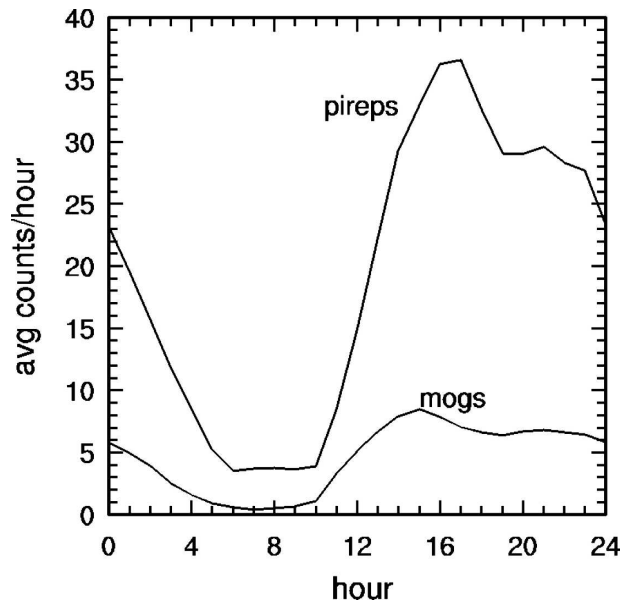


FIG. 2. Diurnal variation of the average number of total and MOG turbulence PIREPs at upper levels (above 20 000 ft MSL) derived from the 3-yr period 2002–2004.

the native resolution NWP output at each grid point in the model domain at the current analysis time. Most of the current diagnostics used are intended to diagnose regions of high turbulence potential due to the presence of upper-level fronts and jet streams, but some are derived from turbulence theory and therefore should be valid for any turbulence source. The suite of diagnostics selected depends on the overall performance of each diagnostic. In addition, the set of diagnostics is selected to ensure that the indices appropriately represent the variety of atmospheric processes that may be contributing to the existing turbulence conditions (i.e., to ensure that the diagnostics are uncorrelated with each other). In general, the diagnostic performance is highly variable; see, for example, the box plots in Fig. 2 of Tebaldi et al. (2002). We have tried as many as 40 different turbulence diagnostics, but currently use a subset that has demonstrated minimum scatter and therefore the best overall performance. The algorithms in this subset are listed below with appropriate references, and the algorithmic expressions for these and others in the GTG suite are provided in appendix A. At upper levels, the following 10 algorithms are used:

- 1) Colson–Panofsky index (Colson and Panofsky 1965);
- 2) Richardson number (Ri ; e.g., Endlich 1964; Kronebach 1964; Dutton and Panofsky 1970);
- 3) diagnostic turbulent kinetic energy (TKE) formulation (DTF3; Marroquin 1998);

- 4) frontogenesis function (isentropic coordinates; e.g., Bluestein 1993);
- 5) unbalanced flow diagnostic (Knox 1997; McCann 2001; Koch and Caracena 2002);
- 6) horizontal temperature gradient (Buldovskii et al. 1976);
- 7) Turbulence Index 1 (TI1; Ellrod and Knapp 1992);
- 8) North Carolina State University index (NCSU1; Kaplan et al. 2004);
- 9) structure function–derived eddy dissipation rate (EDR; Frehlich and Sharman 2004a; Lindborg 1999); and
- 10) structure function–derived sigma vertical velocity (SIGW; Frehlich and Sharman 2004b).

And at midlevels the nine algorithms used are

- 1) TI1 (Ellrod and Knapp 1992);
- 2) wind speed \times horizontal deformation (Reap 1996);
- 3) Absolute value “inertial advection – centrifugal wind” (ABSIA; McCann 2001);
- 4) horizontal temperature gradient (Buldovskii et al. 1976);
- 5) wind speed (e.g., Endlich 1964);
- 6) NCSU1 (Kaplan et al. 2004);
- 7) structure function–derived EDR (Frehlich and Sharman 2004a; Lindborg 1999);
- 8) structure function–derived SIGW (Frehlich and Sharman 2004b); and
- 9) frontogenesis function (pressure coordinates) (e.g., Bluestein 1993).

b. Step 2

In step 2 D_n is interpolated to common flight levels (in increments of 1000 ft) and mapped to a common turbulence intensity scale $0 \leq D_n^* \leq 1$, where 0 corresponds to no turbulence (null) and 1 corresponds to extreme turbulence. This same scale is also used for PIREP intensities to allow quantitative comparisons. A required input for combining the various turbulence diagnostics is the set of threshold values that distinguish the null-light, light-moderate, moderate-severe, and severe-extreme turbulence categories. These thresholds are derived by comparing the PIREP values with the index values for many index–PIREP pairs (essentially a climatology), and computing the median index value corresponding to each turbulence intensity category. The median values of null, light, moderate, severe, and extreme are in turn associated with values of 0, 0.25, 0.5, 0.75, and 1.0, respectively, on the common turbulence intensity scale. The mapping process is performed using a piecewise linear function as shown schematically in Fig. 3. The breakpoints at which the derivative (slope) of the function changes are the thresholds and are given

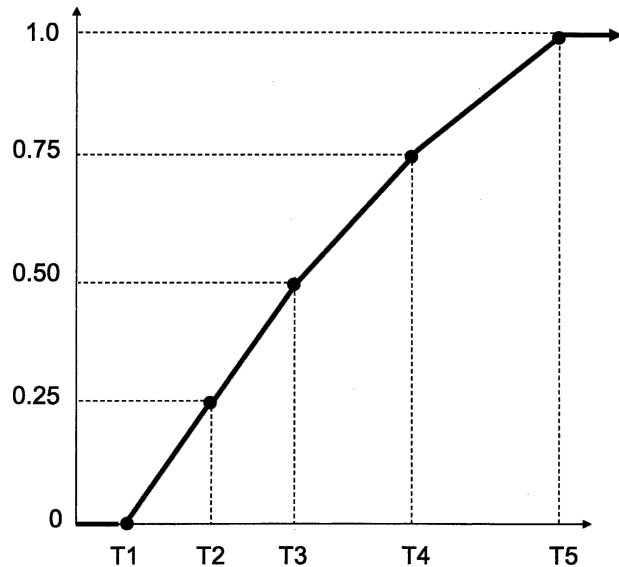


FIG. 3. An index mapping diagram where the raw values of the index are read along the abscissa, with specified thresholds T1 corresponding to the index value for null, T2 for light, T3 for moderate, T4 for severe, and T5 for extreme turbulence values. These are mapped to a 0–1 scale as indicated on the ordinate, with 0.25 being the light, 0.5 the moderate, 0.75 the severe, and 1.0 the extreme threshold. Note that raw index values $<T1$ are always mapped to null, and raw index values $>T5$ are always mapped to extreme. The threshold values T1–T5 are given in Table B1 for each turbulence index.

in appendix B for each index in the current GTG suite. Linear interpolation is performed within each range.

c. Step 3

When using the dynamic weighting strategy, each diagnostic is compared with the available observations (PIREPs) within a time window (currently ± 90 min) around the current NWP model time. For each altitude band of interest, a “score” is determined that measures the relative error between the turbulence intensity as predicted by each diagnostic and the available turbulence PIREPs. There are a number of options available for scoring, but based on previous verification studies of icing and turbulence (e.g., Brown et al. 1997, 2000; Tebaldi et al. 2002), a particularly robust method is to score using probabilities of detection. In this method a contingency table of observations (PIREPs) and forecasted turbulence index values is formed and a PODY, representing the probability of detection of a moderate-or-greater (MOG) event (“yes”), and a PODN, representing the probability of detection of a null or smooth event (“no”), are computed. As shown by Brown et al. (1997) and Brown and Young (2000), combinations of PODY and PODN are preferable to the use of the false

alarm ratio in assessing statistical performance since they are less susceptible to the relative frequencies of yes and no PIREPs, that is, reporting biases.

Another important performance metric is the MOG-forecasted volume occupied by the diagnostic. Based on various atmospheric sampling programs, small MOG volumes are expected at any given time. An example is given in Fig. 4. This is a plot of the distribution of binned eddy dissipation rate (actually $\varepsilon^{1/3}$) automated measurements (Cornman et al. 1995) from about 85 United Airlines (UAL) B757 aircraft in cruise collected over a 3-month period. Note that 99.6% of the measurements fall in the first bin. Since this bin contains what corresponds to both null and light turbulence PIREPs, the fraction of the atmosphere at upper levels containing MOG turbulence should be 1% at most. These percentages are consistent with the results of Dutton (1980).

Therefore, an appropriate metric for evaluating the overall skill of each turbulence diagnostic should include PODY, PODN, and the fraction of the grid volume occupied by MOG turbulence forecasts compared with the computational grid volume f_{MOG} . Similar arguments were made by Brown et al. (1997) for evaluating icing forecasts. A simple scoring function that involves these quantities is

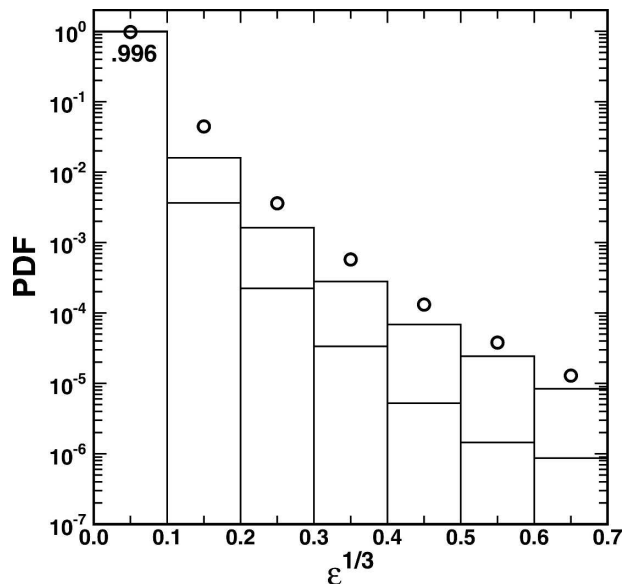


FIG. 4. Distribution of binned $\varepsilon^{1/3}$ median (lower bar) and peak, i.e., 95th percentile (upper bar) values from UAL B757 aircraft over a 3-month time period using the Cornman et al. (1995) algorithm. The open circles are estimates of the distribution based on an assumed lognormal distribution with parameters derived from the RUC-2 NWP model (Frehlich and Sharman 2004a). The difference may reflect the ability of commercial air carriers to successfully avoid turbulence.

$$\varphi = \left(\frac{1.1 + \text{TSS}}{1 + Cf_{\text{MOG}}^p} \right), \quad (1)$$

where the constants C and p can be used to adjust the relative importance of f_{MOG} . In the results to be presented here, $C = 1$ and $p = 0.25$. In the numerator of Eq. (1), the true skill statistic $\text{TSS} = \text{PODY} + \text{PODN} - 1$, and ranges in value from -1 to $+1$, with $+1$ indicating the diagnostic has perfect skill at classifying yes and no PIREPs, and values less than 0 indicating negative skill (0 represents no skill). The TSS is computed by comparing the maximum value of the turbulence diagnostic values D_n^* at the four grid points surrounding each PIREP. Using the average of the four points produces similar results. The thresholds used in evaluating the TSS are given in appendix B (see Table B1). However, not all PIREPs available during the scoring time period are actually used. The goal of the current GTG is to predict turbulence not related to convection. Therefore, a turbulence PIREP that is convectively related should not be used for scoring. Since a PIREP does not specify the source of turbulence, convectively induced turbulence encounters must be identified by indirect means. This is accomplished by comparing each PIREP location with cloud-to-ground (CG) lightning flash data from the National Lightning Detection Network. If the PIREP is within a certain radial distance of a CG lightning flash and is within a certain time window (currently 50 km and 40 min), it is discarded and not used for scoring.

d. Step 4

Once the scoring function φ_n has been computed from step 3, a set of weights W_n can be formed from each diagnostic n :

$$W_n = \left(\frac{\varphi_n}{\sum \varphi_n} \right)^2. \quad (2)$$

Because the number of PIREPs available at any given time is still a small number (for a three hour time window this may be anywhere from less than 10 to as many as 250 depending on the situation), it is not possible to form weights regionally or vertically, so the weights assigned are constant throughout the domain of interest.

e. Step 5

The (dynamically or statically) weighted diagnostics are combined to form the GTG combination. At this point all the diagnostics have been computed and remapped to the 0–1 scale at the initialization time at each grid point (i, j, k) . The GTG diagnostic is now

computed for the initialization time as the simple weighted sum of the diagnostics:

$$\text{GTG}(i, j, k) = \sum_n W_n D_n^*(i, j, k). \quad (3)$$

f. Step 6

The GTG forecasts are formed. When using the dynamic weighting method, an assumption must be made about the temporal variability of the weights. The simplest assumption is that they are constant in time throughout the maximum forecast duration (12 h for RUC-2). This essentially is equivalent to assuming that the state of the atmosphere responsible for turbulence generation processes persists for this length of time. This may be on the long side; Vinnichenko et al. (1980) estimate that the probability of a turbulence patch persisting for longer than 6 h is no more than 50%. In the current GTGD implementation though, the weights are reevaluated every major RUC update cycle, that is, every 3 h, so that the change in large-scale behavior is at least captured at these intervals. In addition, advection (and perhaps growth and decay) of turbulence generation zones should be handled to some extent by the NWP forecast.

Assuming the weights are constant for a given model run, the GTGD forecast procedure is simple. The NWP gridded data are obtained for all *forecast* times (3, 6, 9, and 12 h), the same set of diagnostics is computed from the forecast fields, and the weights derived from the *analysis* time are used in Eq. (3) to get the GTG forecast.

Currently, the entire cycle repeats with every major NWP update; for RUC-2 this is every 3 h. The process is performed separately for midlevels and upper levels, and the results are merged at the FL200 boundary. This was necessary since it was found that

- 1) the best set of turbulence diagnostics was not the same at midlevels and upper levels,
- 2) their optimum threshold values were not the same, and
- 3) the number of available PIREPs was substantially less at midlevels than at upper levels.

When using climatological weights within GTGC, steps 3 and 4 are bypassed and a constant set of weights are used for the analysis time and all forecast times. The procedure for deriving the climatological or default weights is given in the next section and their current values are provided in appendix B.

3. GTG2 performance statistics

In this section various performance statistics are provided for GTG and its component diagnostics. The

computed performance is based on inputs from the RUC-2 NWP model, in its 2003 configuration (roughly 20-km horizontal resolution and 50 vertical levels). Alternative methods of combining the individual diagnostics are also examined, and sensitivity studies are provided to estimate the effects of uncertainties in the verification data sources. Here, performance statistics are derived from comparisons with the only routine observations of aircraft-scale atmospheric turbulence available: verbal reports of turbulence by pilots, or PIREPs. This verification source, although not ideal, has been used in other aviation weather verification studies of both icing (e.g., Brown et al. 1997) and turbulence (Brown et al. 2000; Tebaldi et al. 2002; Lee et al. 2003).

a. PIREPs

PIREPs provide information about a turbulence encounter (time, latitude, longitude, altitude, and severity). A fairly comprehensive review of PIREP reporting and dissemination practices is given in Schwartz (1996). The PIREPs used in this study are received through the NWS's Family of Services communication gateway (see <http://www.nws.noaa.gov/datamgmt/fos/fospage.html> for a description) and augmented by proprietary reports from two major airlines. The raw textual PIREPs are decoded automatically with some data checking to remove reports with one or more invalid parameters and to discard duplicates. These duplicate and bad data records were a very small percentage of the total (<1%). It should be noted that, because of pilot reporting and recording biases, the distribution of PIREP intensities is not what would be expected from Fig. 4; we find the reported intensity distribution is about 55% null, 27% light, 17% moderate, and 1% severe based on 12 yr of turbulence PIREPs between FL100 and FL450.

As noted by Schwartz (1996), PIREP inaccuracies in time, position, and intensity can lead to some uncertainty in the verification results that will be quantified here to the extent possible when presenting the final results. Further, it must be realized that a report is based on a turbulence experience along a flight path, that is, along a line, but is usually reported as a single point value. If the model-derived diagnostics are supposed to be a grid volume average, the correspondence to a line is not necessarily direct.

Position and time uncertainties in PIREPs were evaluated by comparing the locations of positive UAL in situ turbulence measurements using the automated Cornman et al. (1995) algorithm with PIREPs from the same aircraft. We found that, based on about 450 comparisons over a 4-month period, the median uncertainty was about 50 km horizontally, 200 s in time, and 70 m

vertically. These time and vertical position differences are well within the windows used for verification. The horizontal position uncertainty of 50 km corresponds to about two to three RUC-2 grid points, and since the resulting turbulence forecast fields are fairly smooth both horizontally (cf. Fig. 1) and vertically, this uncertainty should have a negligible effect on the results.

To assess the uncertainty in PIREP turbulence intensities, the intensity values reported by two aircraft close together in space and time (using the 50-km uncertainty found above, at the same flight level, and within a time window of 600 s) were compared using 12 yr of PIREP data at the flight levels assessed in this study (FL100–FL200 and FL200–FL450). For the upper-altitude band this provided about 8200 pairs for comparison, with the results that roughly 84%, 68%, 85%, and 86% agree on intensities of smooth, light, moderate, and severe, respectively, regardless of aircraft weight class. For the midlevel altitude band about 3600 pairs were compared giving percentage agreements of 64%, 46%, 75%, and 77%, respectively, for the major intensity categories. The larger discrepancies at midlevels are probably because of the greater mix of aircraft types at lower flight levels. The relative uncertainty in the light intensity values is understandable given the wide range of aircraft sizes and weights reporting, and it is for this reason they are not used for verification. However, at least statistically, turbulence PIREPs seem to have acceptable position and timing errors, and the null and MOG reports are very consistent. These numbers bolster our confidence in the use of PIREPs for verification, and provide a means for assessing the uncertainty in the results to be presented.

b. GTG2 performance

In previous studies that used PIREPs for both icing (e.g., Brown et al. 1997) and turbulence (e.g., Brown et al. 2000; Tebaldi et al. 2002; Lee et al. 2003) verification, one metric used to evaluate performance was the area contained under the PODY–PODN curves, similar to receiver (or relative) operating characteristic (ROC) curves. In this procedure a set of thresholds is assumed for each diagnostic, and for each threshold the diagnostic performance based on comparisons with available turbulence PIREPs is evaluated by computing both a PODN and a PODY. These curves essentially measure the ability of a forecast algorithm to discriminate between yes and no observations. For small values of the chosen threshold, PODY will be high, near unity, while PODN will be low, near 0, and vice versa for large values of the chosen threshold. For the range of thresholds selected, higher combinations of PODY and PODN, and therefore larger areas under the PODY–

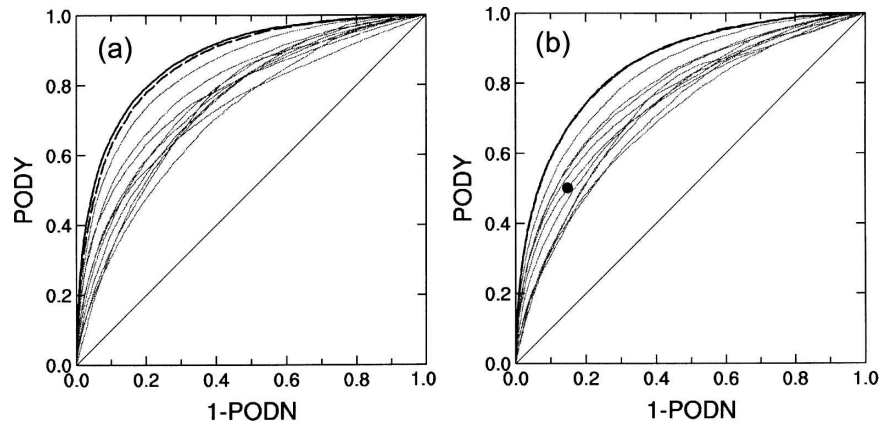


FIG. 5. Individual diagnostics and the GTG2 PODY-PODN performance statistics (individual diagnostics as thin gray, GTG combination as heavy black solid, and GTG combination using climatological weights as heavy black dashed) derived from 1 yr (2003) of (a) 1800 UTC analyses (0-h forecasts) using 37 878 PIREPs and (b) 1800 UTC 6-h forecasts (valid 0000 UTC) using 49 703 PIREPs, for upper levels (FL200-FL460). For comparison, the no skill line is also shown as the diagonal line, and the 2003 average AIRMET performance (with amendments) at upper levels centered on 2100 UTC is shown as a heavy dot in (b).

PODN curves, imply greater skill in discriminating between null and MOG turbulence events. The area under the curve (or AUC) ranges from 0.5 for no skill to 1.0 for perfect skill. For a more complete discussion of the use of the AUC as a discrimination metric, see, for example, Hanley and McNeil (1982), Mason (1982), Kharin and Zwiers (2003), and Marzban (2004).

Samples of PODY-PODN statistical performance derived for GTG2 as described above are provided in Fig. 5 (for upper levels) and Fig. 6 (for midlevels). The curves are based on 1 yr (2003) of PIREP comparisons for 0-h analyses and 6-h forecasts. The 0-h results give some indication of the ability of the individual turbulence diagnostics and the GTG combinations to account for the observed turbulent state of the atmosphere. The 6-h forecast was chosen for assessment because it is an adequate lead time for route planning purposes for almost all continental U.S. (CONUS) flights. All 0-h analyses are taken at 1800 UTC with the GTGD weights formed based on the performance of the individual diagnostics computed from the RUC 1800 UTC analysis. The 6-h forecasts are derived from diagnostics computed from the 6-h RUC forecast, initialized at 1800 UTC, valid at 0000 UTC the next day, and with weights provided from the 1800 UTC analysis (GTGD). These times were chosen because both times correspond to daylight hours over the CONUS where air traffic density is sufficient to provide large numbers of PIREPs for both initialization and verification.

Figures 5 and 6 demonstrate that both the GTGD and GTGC combinations are superior to the individual

diagnostics as measured by the ROC AUC metric. As expected, the AUC is higher for the analysis time (GTGC AUC is 0.878 for upper levels and 0.818 for midlevels) than for the 6-h forecasts (0.852 and 0.792, respectively). For both times the upper-level performance is slightly better than the midlevel performance. This is probably because of the fewer PIREPs available to fit at midlevels and to the experience derived from the GTG1 development, which has allowed formulation of better turbulence diagnostics for upper levels.

For comparison, the AIRMET performance is also shown. AIRMETs are the operational forecasts of turbulence produced by the AWC every 6 h and are valid for up to 6 h (refer to http://aviationweather.gov/exp/product_overlay/help/p-airmets.html for a description) but may be amended as needed between the standard issue times. They are textual products that describe as three-dimensional polygons the regions of forecasted turbulence. Since the polygons are necessarily relatively simple, and are assumed valid for the entire 6 h (although we did allow for amendments), the comparison here is not exact; nevertheless, AIRMETs are the current operational product, and some comparisons must be made to assess the ability of automated forecasts to provide benefit to the aviation community. The AIRMET performance (with amendments) was retrieved from the RTVS archives for the same time window (centered at 2100 UTC) as the model-based algorithms. They are competitive with some of the individual diagnostics, but are not as good as either of the GTG combinations by this measure. Comparisons with

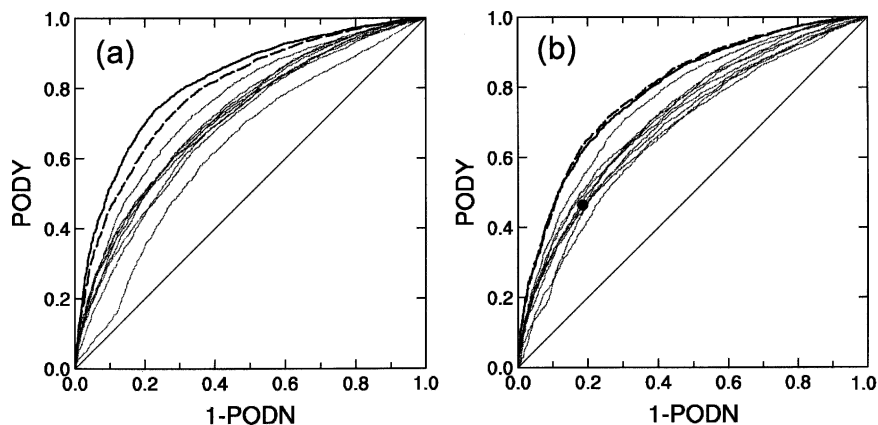


FIG. 6. PODY-PODN curves as in Fig. 5 but for midlevels (FL100-FL200): (a) 1800 UTC analyses (0-h forecasts) using 6575 PIREPs and (b) 1800 UTC 6-h forecasts (valid 0000 UTC) using 8063 PIREPs.

turbulence SIGMETs were not attempted because they are usually based on observations of severe turbulence, most of which tend to be related to convection.

For GTGC, consistent with Eq. (2), the static weights are proportional to the ROC AUC squared for each diagnostic. Note from Figs. 5 and 6 that the use of cli-

matological weights degrades the performance only slightly for the analysis time, and is nearly identical in performance for the 6-h forecast.

An equivalent representation of the relative performance of the individual diagnostics and the GTGD combination is provided by plots of the probability den-

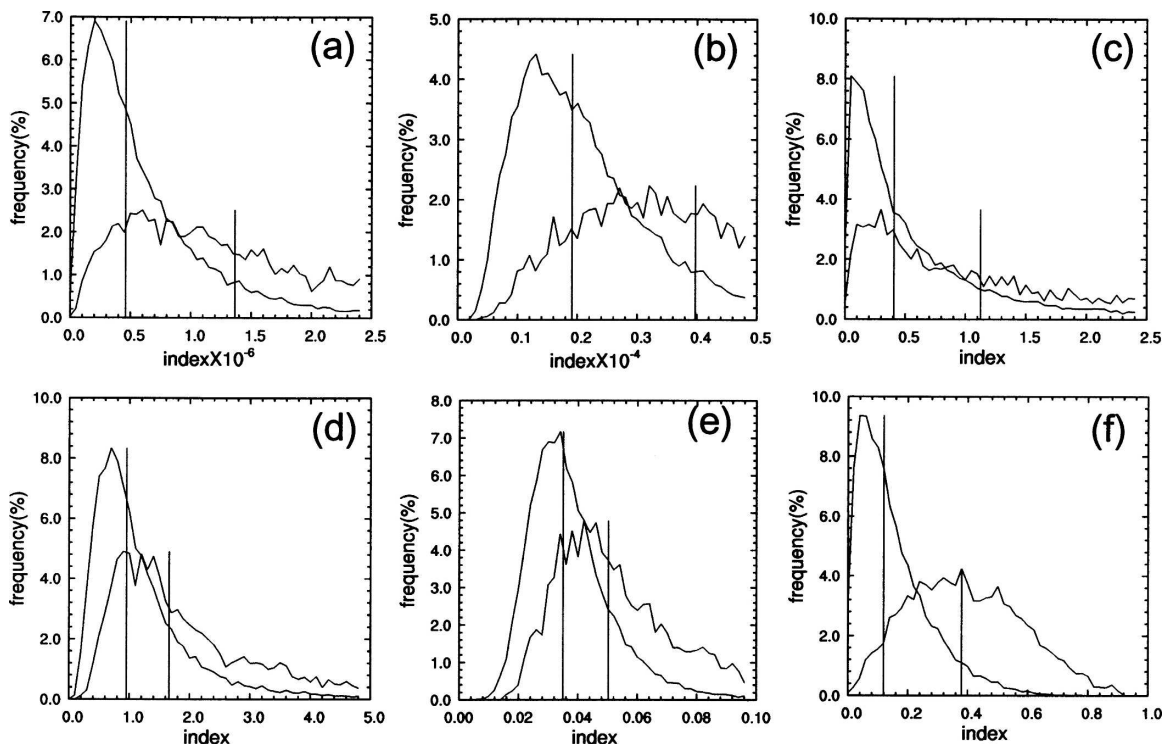


FIG. 7. Probability density curves for five different diagnostics within the GTG combination: (a) TI1, Eq. (A15); (b) horizontal temperature gradient, Eq. (A23); (c) DTF3, Eq. (A6); (d) divergence tendency (UBF), Eq. (A30); (e) eddy dissipation rate, Eq. (A7); and (f) the GTG combination based on 6-h upper-level forecast index-PIREP correlations. In each panel the furthest left curve provides the null distribution and the furthest right curve provides the MOG distribution, and the curves have been normalized to contain the same area. The vertical lines place the medians of the distributions.

TABLE 1. Average 6-h forecast performance metrics for the individual diagnostics and the GTG combination derived from 1 yr (2003) of 1800 UTC 6-h forecasts (valid 0000 UTC). The PODY, PODN, and TSS values are computed using the thresholds given in appendix B. The column labeled “Eq.” refers to the equation number in appendix A.

Index name	Eq.	Rmse/PIREP	PODY	PODN	TSS	f_{MOG}	φ	Avg wt
Upper levels (49 703 PIREPs)								
–Ri	(A1)	0.0643	0.591	0.832	0.423	0.180	0.782	0.101
CP	(A4)	0.0643	0.370	0.884	0.254	0.128	0.689	0.088
DTF3	(A6)	0.0761	0.452	0.824	0.276	0.078	0.751	0.100
EDR	(A7)	0.0620	0.440	0.872	0.311	0.068	0.845	0.106
SIGW	(A8)	0.0863	0.525	0.788	0.313	0.188	0.738	0.090
F_θ	(A9)	0.0525	0.548	0.889	0.437	0.095	0.882	0.119
TI1	(A15)	0.0672	0.589	0.846	0.434	0.104	0.878	0.113
$ \nabla_H T $	(A23)	0.0735	0.633	0.798	0.431	0.199	0.809	0.095
UBF	(A30)	0.0687	0.372	0.865	0.237	0.110	0.726	0.085
NCSU1	(A36)	0.0761	0.536	0.831	0.367	0.101	0.832	0.103
GTGD	(3)	0.0397	0.530	0.926	0.456	0.069	0.968	—
GTGC	(3)	0.0397	0.527	0.926	0.453	0.071	0.961	—
AIRMETS	—	—	0.501	0.852	0.353	0.134	0.905	—
Midlevels (8063 PIREPs)								
EDR	(A7)	0.1189	0.619	0.714	0.333	0.109	0.812	0.119
SIGW	(A8)	0.1444	0.660	0.655	0.315	0.170	0.761	0.107
F_p	(A10)	0.0807	0.455	0.829	0.284	0.058	0.848	0.124
TI1	(A15)	0.0891	0.534	0.770	0.304	0.104	0.781	0.114
$ \nabla_H T $	(A23)	0.1096	0.570	0.710	0.280	0.133	0.724	0.102
$ \mathbf{v} $	(A24)	0.1270	0.514	0.746	0.260	0.180	0.648	0.096
$ \mathbf{v} \text{ DEF}$	(A28)	0.0985	0.581	0.787	0.369	0.110	0.817	0.116
ABSIA	(A34)	0.0913	0.580	0.744	0.324	0.110	0.784	0.110
NCSU1	(A36)	0.0858	0.380	0.846	0.226	0.057	0.819	0.112
GTGD	(3)	0.0541	0.614	0.813	0.439	0.078	0.956	—
GTGC	(3)	0.0528	0.623	0.816	0.427	0.079	0.961	—
AIRMETS	—	—	0.464	0.815	0.279	0.083	0.8971	—

sity functions (PDFs) for both null and MOG turbulence encounters. These are shown in Fig. 7. A perfect diagnostic would have no overlap between the null and MOG PDF curves, and so the amount of overlap is a measure of the PODs. Qualitatively, there is substantial overlap for all indices, but the overlap is clearly minimized with the GTG combination, reinforcing the overall robust nature of the GTG approach.

Other performance measures are shown in Table 1, namely, the 6-h forecast average PODN, PODY, TSS, root-mean-square error (rmse) per PIREP, f_{MOG} , φ , and dynamic weight for each turbulence diagnostic; the GTG combinations; and AIRMETS. By almost any measure, the GTG combinations provide superior performance for both mid- and upper levels. From the average weights given in Table 1, the single best diagnostic is the frontogenesis function [Eq. (A9)] evaluated in isentropic coordinates at upper levels and evaluated on constant pressure surfaces [Eq. (A10)] at midlevels. Consistent with Figs. 5 and 6 the AIRMET performance has similar skill to some of the individual diagnostics in terms of both statistical error and fraction of MOG airspace forecasted.

c. Sensitivity studies

As stated earlier there are reporting biases, as well as position, timing, and intensity errors associated with PIREPs, consequently the verification performance statistics are subject to some amount of uncertainty. To attempt to quantify this uncertainty, we have performed two sensitivity studies. The first study addresses the irregular nature of the PIREPs distribution (in frequency and location) by degrading the quality of the data distribution by randomly resampling only a fraction of the available PIREPs and using that fraction for verification. Specifically, from the full set of PIREP–GTG forecast data pairs, five subsets of 1/2, 1/3, 1/4, 1/5, and 1/6 of the available pairs were used for scoring. Two hundred subsets were used, and for each subset a ROC curve was computed. The solid curves in Fig. 8 show the derived envelope of “uncertainty” around the original GTG ROC 6-h forecast curves. The results plotted are for GTGC, but identical results were obtained with GTGD. The uncertainty bands are very tight for high levels (about $\pm 2\%$), but are somewhat wider for midlevels (about $\pm 6\%$) because of the

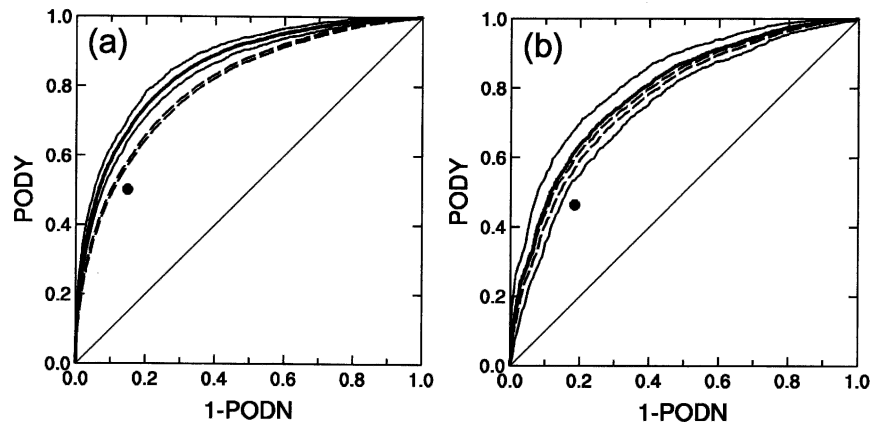


FIG. 8. GTGC PODY-PODN plots for 6-h forecasts at (a) upper levels and (b) midlevels accounting for two measures of uncertainty. The heavy solid curve in both figures uses the entire set of verification PIREPs unperturbed in intensity (cf. Figs. 5b and 6b). The thin dashed lines represent the uncertainty bounds obtained by resampling 200 subsets of the available PIREPs. The dashed lines represent the uncertainty bounds obtained by randomly modifying 25% of the PIREP intensities by one full intensity category and repeating 200 times.

smaller number of PIREPs available at midlevels. Note that the lower bound of the uncertainty curve still provides significantly better PODY-PODN performance than AIRMETs. A comparison with the best single index was also performed (not shown) by computing the difference between these curves for each subset of PIREPs at the highest point on the curves corresponding to the best TSS value, across the 200 randomized subsets. The 95% confidence interval for this difference (GTG-best index) is always positive for both PODN and PODY. These results are consistent with those of Kane and Brown (2000) using an earlier version of GTG.

The second study addresses the uncertainty in the reported intensity values of the PIREPs. This problem

was discussed in section 3a, and there it was stated that, climatologically, the percentage agreement in intensities reported by nearby aircraft was somewhere between 70% and 80% on average. Based on these results, a “perturbation” experiment was performed by assuming that 25% of the verification PIREPS may be incorrectly reported by one full intensity category (e.g., from light to moderate). The sensitivity to PIREP intensity uncertainty was then assessed by either randomly increasing or decreasing by one intensity level 25% of the available verification PIREPs. The process was repeated 200 times for different subsets of the verification PIREPS.

The dashed curves in Fig. 9 show the results in terms of an envelope of “downgraded” ROC curves. Note

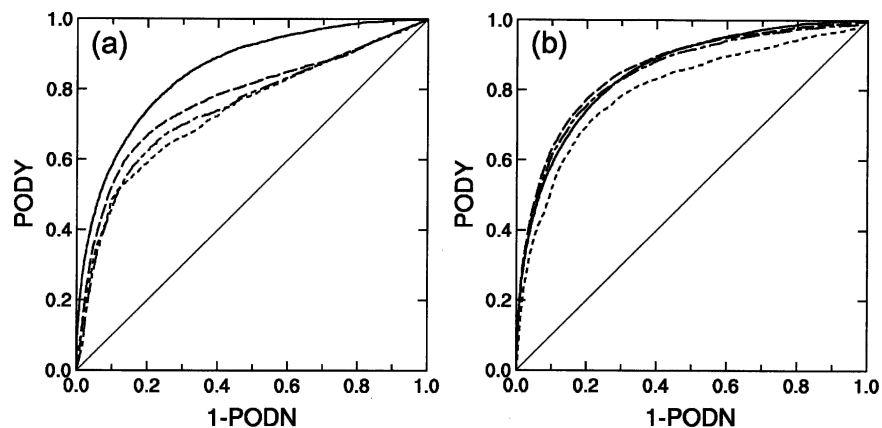


FIG. 9. PODY-PODN plots for the 6-h upper-level forecasts as in Fig. 5 but comparing four different regression strategies for combining the diagnostics: GTG (solid), logistic (dashed), decision tree (dotted), and neural network (dash-dot), using a training set corresponding to PIREPs available from (a) one analysis time and (b) 14 analysis times.

that the degradation (from about -5% to -6%) in the upper-level ROC curves is greater than for the midlevel ROC curves (from about -1% to -3%). This may be due in part to the greater inherent uncertainty in PIREP intensities at midlevels noted above in section 3a, and in part by the finer GTG tuning achieved at upper levels compared with midlevels making the upper-level results more sensitive to degradation. Still the ROC curve for GTG significantly outperforms the unperturbed AIRMETS and the single indices' curves (not shown), as confirmed through the computation of the differences as described above in the resampling exercise.

This is a rather severe test; when the process was repeated by perturbing the PIREPs by only a one-half intensity category (e.g., from light to light-moderate) the resulting GTG ROC curves were almost indistinguishable from the original.

d. Other combinatorial methods

Other methods for combining the indices were also tested. In particular, three well-established statistical models for forecasting a binary outcome (with null encounters of turbulence being 0s, and MOG encounters being 1s) were investigated and their predictive performance compared with the GTG method:

- (a) logistic regression,
- (b) tree classification, and
- (c) neural networks.

Logistic regression fits a linear combination of the diagnostics to the probability of a positive (i.e., MOG) outcome. For a description of the method, see, for example, McCullagh and Nelder (1989). The estimation is likelihood based, so that the optimal solution for the coefficients of the linear combination maximizes the probability of the training set of data, under the assumption of a binomial distribution; namely,

$$P(\text{PIREP} = \text{MOG} | D_1, D_2, \dots, D_N) = \frac{\exp\left(\beta_0 + \sum_{n=1}^N \beta_n + D_n\right)}{1 + \exp\left(\beta_0 + \sum_{n=1}^N \beta_n D_n\right)},$$

where D_n is the value of diagnostic n matched to each PIREP and β_n is the derived coefficient for D_n .

Classification based on recursive binary partitions (tree classification) is another popular model for predictions (see, e.g., Breiman et al. 1984). Starting with the full set of observation–individual diagnostic “pairs,” the algorithm looks at every diagnostic in turn, in order to determine an optimal value for splitting the dataset.

The value must be such that the two subsets of PIREPs thus created (PIREPs matched to values of the diagnostic less than the splitting threshold, and PIREPs matched to values of the diagnostic greater than the splitting threshold) are as homogeneous as possible within and as heterogeneous as possible between. Ideally, if a diagnostic were a perfect discriminator of null versus MOG turbulence encounters, a value of that diagnostic would exist that would separate the PIREP set into two groups perfectly homogeneous “within” (all nulls on one side, all MOGs on the other) and thus perfectly heterogeneous “between.” After the first split is performed, the algorithm is applied independently to the two groups of observations thus formed. Each split chooses a diagnostic and a value within the diagnostic range. The algorithm may separate the data perfectly by splitting the groups until only one observation is left in each of the “leaves” of the tree. However the splitting rules thus defined are bound to be too ad hoc with respect to the training set, and would certainly constitute a bad model for forecasting purposes. Thus, tree classification algorithms must be optimized in order to achieve a balance between fitting the training set of data and performing accurate prediction on an independent set.

Neural networks are flexible regression methods that substitute a nonlinear combination of diagnostics for the linear form assumed by logistic models. Similar to tree models, a balance has to be found between extremely flexible structures that are able to fit the training set almost perfectly, and a more parsimonious representation that has better generalization ability. For a more complete discussion, see, for example, Ripley (1996).

Each of the three statistical models was executed using the R, version 2.0 (see Ihaka and Gentleman 1996 for a description), statistical analysis package (routines glm, rpart, and nnet for logistic regression, decision trees, and neural nets, respectively). The performance for each is based on the same set of observations as used to obtain the GTG results. PIREP–diagnostic “pairs” from the analysis time were used to train the models, and then evaluated on the 6-h forecast. As the curves in Fig. 9a demonstrate, GTG does outperform the estimates derived from all three competing methods. This is probably because of the relative scarcity of the training data available, which is insufficient for robust estimations using highly multivariate models.

The performances become very close (almost indistinguishable for logistic and neural networks, still subpar for tree regression) as the training set is increased to include 14 days of PIREP–diagnostic pairs, as Fig. 9b shows. By increasing the number of cases in the training

set, the estimates of the many parameters in the statistical models stabilize and are not so susceptible to small, nonrepresentative (at least in terms of the functional forms that are being fitted) groups of observations. Evidently, the functional form used in the GTG optimization procedure is simple enough, and tuned explicitly to the POD performance measure, to allow an effective estimation on the basis of a single day of training, thus offering an extremely manageable solution to the operational implementation of the algorithm. Obviously, it would be much more expensive from a storage and computational perspective if a longer set of training data had to be processed at every forecast issue time.

4. Summary and conclusions

In summary, the overall performance of GTG seems to be skillful enough to provide useful information to meteorologists and dispatchers for strategic planning for turbulence avoidance. In particular the following points have been shown.

- The GTG combination provides superior PODY-PODN performance over that available by using a single turbulence diagnostic or by AIRMETS.
- Although the dynamic weighting method gives the best nowcast performance, the use of a set of static climatologically derived weights provides forecast performance comparable to the dynamic weighting method, making it an attractive alternative for use in data-sparse (PIREPs) regions.
- The simpler weighting strategies used within the GTG framework provide performance comparable to more complicated procedures, such as neural networks, and generally requires less “training.”
- In this investigation, based on the average dynamic weights chosen by GTGD (see Table 1) the single best diagnostic at both midlevels and upper levels is the frontogenesis function.

At this point a few caveats should be stated. The results presented here are based on inputs from the RUC-2 NWP model, in its 2003 configuration (roughly 20-km horizontal resolution with 50 vertical levels). The use of a different NWP model and/or different vertical or horizontal resolutions may be expected to change some of these results, including the relative performance of the individual turbulence diagnostics and the GTG combinations. Further, the analyses are based on the performance derived from an entire year of data over the entire RUC-2 computational domain. Daily, seasonal, and regional dependencies would be expected in the performance statistics.

The ability to provide still more accurate aircraft-

scale turbulence nowcasts and forecasts is hampered by several fundamental difficulties. First, the resolution of current NWP models (several 10s of km) is about two orders of magnitude too coarse to resolve aircraft-scale turbulence (roughly 100s of m). Therefore, aircraft-scale turbulence diagnoses and predictions must be based on resolvable- (by the NWP) scale features. However, and this is the second difficulty, the performance of turbulence diagnostics is hampered by our current lack of understanding of the linkage between NWP resolvable-scale features and aircraft-scale turbulence. An implicit assumption underlying the use of all these diagnostics is that turbulence-generating mechanisms have their origin at resolvable scales and that energy cascades down to aircraft scales, but it is unclear what the exact cascade mechanism is. Recent high-resolution simulations by Lane et al. (2004, 2005) indicate that the linkage is related at least in some cases to gravity wave production by features such as upper-level fronts and convection, and subsequent breakdown of the waves into turbulence. Third, even if it is true that aircraft-scale turbulence has its origins at the resolvable scales, the turbulence forecast system has all the inherent NWP errors associated with the resolvable scales. Fourth, it is not clear that the current suite of turbulence diagnostics is in fact capturing all the relevant information that the larger-scale representations can provide.

Then there is the difficult matter of model fitting and verification. The GTG system uses PIREPs for tuning and verification. But as shown, an individual PIREP is subject to spatial, temporal, and intensity misrepresentations. The quantitative automated in situ turbulence reporting system (Cornman et al. 1995, 2004) should eliminate most of the uncertainty associated with PIREPs but will still not alleviate the nighttime underreporting bias. The advantages of this data for GTG are obvious. First, the data will be more accurate than PIREPs, for both intensity and position. Second, the amount of data will be vastly increased, since the current plan is to relay turbulence every minute in cruise. This will provide a much more complete mapping of the turbulent state of the atmosphere (at least at upper levels), and will allow GTG to fit that state much more precisely than has been possible using the current set of scattered PIREPs. Just as the accuracy of upper-level winds in NWP models has increased with the use of Aircraft Communications Addressing and Reporting System wind data (e.g., Schwartz et al. 2000), the GTG upper-level forecasts should become more accurate with the ingest of in situ data.

Efforts continue to provide a better turbulence forecasting system through the following research areas.

- Better diagnostics—this is a continued research area at major laboratories and universities. But any diagnostic must be judged by its overall performance, not just on a few select cases. In addition, information about when a particular diagnostic performs well and when it does not could be used in dynamically assigning its weight within the GTG framework. Hopkins (1977) and Lester (1994) describe synoptic conditions that are known to be conducive to CAT, and these could be developed into automated algorithms. Although some of the current diagnostics are independent of the source of turbulence (e.g., Ri), most are tuned for CAT associated with upper-level fronts and enhanced wind shears in the vicinity of jet streams. Diagnostics for other known sources of turbulence, for example, those related to deep convection, must be developed, tested, and implemented into future versions of GTG. Turbulence related to convection has been shown by Kaplan et al. (2005) to be coincident with some particularly severe turbulence encounters. Nevertheless, the current version of GTG does capture some turbulence cases related to convection, if the convection is associated with mid- to upper-level disturbances that may be identified by one or more of the current diagnostics.
- “Local” fits—within the current GTG framework, the best fit of diagnostics is determined for the entire volume of atmosphere between the altitude bands of interest. Better fits are probably attainable in subvolumes that could be overlapped to give smooth transitions from one subvolume to another. Although the number of PIREPs available for regional or local fits is probably insufficient at the current time, the use of the turbulence in situ measurements may allow for local fits, both horizontally and vertically.

Acknowledgments. We appreciate the careful readings and suggestions for improvement of an earlier version of the manuscript by Barbara Brown, Jennifer Abernethy, Teddie Keller, and Rod Frehlich. We also thank the three anonymous reviewers for their comments that helped improve the presentation of the material. We also thank Jennifer Mahoney, NOAA-Research-ESRL/GSD, for supplying the relevant AIRMET data. This research is in response to requirements and funding by the FAA. The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA.

APPENDIX A

GTG Turbulence Diagnostics

This appendix lists the current suite of turbulence diagnostic algorithms within GTG. Although all these

are or have been computed and evaluated, only a (user selectable) subset is actually included in the GTG combination. Note that in some cases the constituent components of a diagnostic may themselves be used as a turbulence index.

a. Richardson number and its components

The Richardson number and its components are well-known turbulence diagnostics (e.g., Endlich 1964; Kronebach 1964; Dutton and Panofsky 1970, etc.). Theory and observations have shown that at least in some situations clear-air turbulence patches are produced by Kelvin–Helmholtz instabilities. This occurs when Ri becomes small. Therefore, theoretically, regions of small Ri should be favored regions of turbulence, where

$$\text{Ri} = \frac{N^2}{S_V^2}, \quad (\text{A1})$$

with

$$N^2 = \frac{g}{\theta} \frac{\partial \theta}{\partial z} \quad \text{or} \quad \frac{g}{\theta_e} \frac{\partial \theta_e}{\partial z} \quad (\text{A2})$$

and

$$S_V = \left| \frac{\partial \mathbf{v}}{\partial z} \right| = \left(\left| \frac{\partial u}{\partial z} \right|^2 + \left| \frac{\partial v}{\partial z} \right|^2 \right)^{1/2}. \quad (\text{A3})$$

Here, θ is potential temperature, θ_e is equivalent potential temperature, g is the acceleration due to gravity, z is the vertical direction, and \mathbf{v} is the horizontal wind vector with components u and v in the east–west and north–south directions, respectively.

b. TKE

TKE formulations are based on the TKE balance equation, assuming horizontal homogeneity and stationarity. The Colson–Panofsky index (CP; Colson and Panofsky 1965) uses dimensional arguments in a stable atmosphere to estimate clear-air turbulence intensities as

$$\text{CP} = \lambda^2 S_V^2 \left(1 - \frac{\text{Ri}}{\text{Ri}_{\text{crit}}} \right), \quad (\text{A4})$$

where λ is a length scale, taken as the local value of vertical grid increment Δz , and Ri_{crit} is an empirical constant (≈ 0.5).

Laikhtman and Al'ter-Zalik (1966; also Vinnichenko et al. 1980) developed a similar equation but with a different prescription for the length scale:

$$\text{TKE} = C\Phi^{3/2}(d\Phi/dz)^{-1}, \quad (\text{A5})$$

where C is an adjustable constant and

$$\Phi = S_V^2 - \alpha N^2 > 0,$$

where $\alpha = 1/\text{Pr} = K_H/K_M$ is taken as another adjustable constant, Pr is the turbulent Prandtl number, and K_H and K_M are the eddy diffusivities of heat and momentum, respectively.

Marroquin (1998) diagnostic TKE formulations (DTF) used k - ε closure equations (e.g., Stull 1988) and other simplifications to derive diagnostics for TKE and/or ε , giving, for example, for DTF3,

$$\varepsilon = K_M \left(\frac{c_1}{c_3} S_V^2 - \frac{c_2}{c_3} \frac{N^2}{\text{Pr}} \right), \quad (\text{A6})$$

where $c_1 = 1.44$, $c_2 = 1.0$, and $c_3 = 1.92$ (Stull 1988, p. 219), and K_M and Pr are taken as adjustable constants to get the best agreement with observations.

c. Eddy dissipation rates

Eddy dissipation rates are estimated from second-order structure functions (Frehlich and Sharman 2004a,b). The one-dimensional structure function of variable $q(x)$ is defined as

$$D_q(s) = \langle [q(x) - q(x+s)]^2 \rangle,$$

where s is the separation distance and $\langle \rangle$ denotes an ensemble average. The structure functions of the velocity components parallel or normal to the displacement vector $\mathbf{s} = (x, y, z)$ can be related to turbulence intensity ε (for $q = u, v$) or σ_w^2 (for $q = w$, the vertical velocity component) through

$$D_q(s) \propto C_q(s) \varepsilon^{2/3} D_{\text{REF}}(s) \quad (\text{A7})$$

and

$$D_w(s) \propto C_w(s) \sigma_w^2, \quad (\text{A8})$$

where $C_q(s)$ and $C_w(s)$ take into account NWP model specific spatial filtering effects and D_{REF} is given by Lindborg (1999); for small separations s it is proportional to $s^{+2/3}$. In the text and tables the relation (A7) to derive $\varepsilon^{1/3}$ is indicated as EDR and relation (A8) to derive σ_w is indicated as SIGW.

d. Frontogenesis function

Fronts contain regions of low Ri and therefore may be conducive to turbulence (e.g., Jacobi et al. 1996) and can also be a source of gravity waves that may be unstable (e.g., Lane et al. 2004). The definition of the frontogenesis function is (e.g., Bluestein 1993, p. 253)

$$F = \frac{D}{Dt} |\nabla\theta|,$$

where D/Dt is the Eulerian time derivative. This can be rewritten in two dimensions, using the thermal wind relation, as

$$F \propto \frac{D}{Dt} \left[\left(\frac{\partial u}{\partial \theta} \right)^2 + \left(\frac{\partial v}{\partial \theta} \right)^2 \right]^{1/2} = \left| \frac{\partial \mathbf{v}}{\partial \theta} \right|^{-1} \left[\frac{\partial u}{\partial \theta} \frac{D}{Dt} \left(\frac{\partial u}{\partial \theta} \right) + \frac{\partial v}{\partial \theta} \frac{D}{Dt} \left(\frac{\partial v}{\partial \theta} \right) \right].$$

Expanding on a constant- θ surface and invoking continuity gives

$$F_\theta \propto - \frac{D}{Dt} \left[\left(\frac{\partial u}{\partial \theta} \right)^2 + \left(\frac{\partial v}{\partial \theta} \right)^2 \right]^{1/2} = \left| \frac{\partial \mathbf{v}}{\partial \theta} \right|^{-1} \left[\frac{\partial u}{\partial \theta} \frac{D}{Dt} \left(\frac{\partial u}{\partial \theta} \right) + \frac{\partial v}{\partial \theta} \frac{D}{Dt} \left(\frac{\partial v}{\partial \theta} \right) \right]. \quad (\text{A9})$$

This is the form used in GTG at upper levels. Note that its formulation is based on an isentropic coordinate system (as used at upper levels in the RUC model). In that case an alternative is to use a constant pressure p surface formulation:

$$F_p \propto \frac{\partial u}{\partial p} \frac{D}{Dt} \left(\frac{\partial u}{\partial p} \right) + \frac{\partial v}{\partial p} \frac{D}{Dt} \left(\frac{\partial v}{\partial p} \right). \quad (\text{A10})$$

This is the form used in GTG at midlevels.

e. Richardson number tendency

The index, $d\text{Ri}/dt$ (Roach 1970; Keller 1990), is based on attempts by several investigators to forecast turbulence by using a time tendency equation for the Richardson number. The version used within GTG is based on a formulation of this equation in isentropic coordinates by Keller (1990), termed Specific CAT Risk (SCATR), given by

$$\text{SCATR} = \frac{1}{24} \Phi S_V^2, \quad (\text{A11})$$

with

$$\Phi = \frac{2}{S_V^2} F_\theta - \Delta_H, \quad (\text{A12})$$

and

$$\Delta_H = \text{horizontal divergence} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}.$$

Brown's index (Brown 1973) is a simplification of the Ri tendency equation originally derived by Roach (1970). The simplifications involve use of the thermal wind relation, the gradient wind as an approximation to the horizontal wind, and some empiricism. Form (A13) is the simplified Ri tendency equation, while form (A14) is an extension to provide a measure of turbulence intensity:

$$\Phi = (0.3 \zeta_a^2 + D_{SH}^2 + D_{ST}^2)^{1/2} \quad (A13)$$

and

$$\varepsilon = \frac{1}{24} \Phi S_V^2, \quad (A14)$$

where the shearing deformation $D_{SH} = \partial v/\partial x + \partial u/\partial y$, the stretching deformation $D_{ST} = \partial u/\partial x - \partial v/\partial y$, absolute vorticity $\zeta_a = \zeta + f$, with $\zeta = \partial v/\partial x - \partial u/\partial y$ and f is the Coriolis frequency.

f. Ellrod indices

Ellrod indices (Ellrod and Knapp 1992) are derived from simplifications of the frontogenetic function. As such it depends mainly upon the magnitudes of the horizontal gradients of θ (proportional to S_V through the thermal wind relation) and the total deformation. The two variants developed were

$$TI1 = S_V DEF \quad (A15)$$

and

$$TI2 = S_V (DEF - \Delta_H), \quad (A16)$$

where

$$DEF = (D_{SH}^2 + D_{ST}^2)^{1/2}. \quad (A17)$$

g. Potential vorticity

Another index is the potential vorticity (PV) (Knox 2001) or horizontal gradient of PV (Shapiro 1978):

$$|PV| \quad (A18)$$

and

$$|\nabla PV|, \quad (A19)$$

where

$$PV = -g \zeta_a \frac{\partial \theta}{\partial p}.$$

h. Clark's CAT algorithm

Clark's CAT (CCAT; Vogel and Sampson 1996) index has been used by U.S. Navy forecasters for at least two decades. It is defined as

$$CCAT = \left(\frac{g}{T}\right) \left(\frac{\zeta_a}{f}\right) \left(\mathbf{v} \cdot \nabla \left| \frac{\partial T}{\partial z} \right| \right), \quad (A20)$$

where T is absolute temperature.

i. Curvature measures

Tight curvatures in upper-level troughs and ridges have been empirically related to turbulence outbreaks (e.g., Lester 1994; Hopkins 1977; Arakawa 1952; Stone 1966). A measure of curvature is relative vorticity ζ , with a maximum (positive value) in upper-level troughs and a minimum (negative value) in upper-level ridges. Therefore,

$$\zeta^2 = |\nabla \times \mathbf{v}|^2 \quad (A21)$$

should be a measure of curvature in both troughs and ridges.

Another measure may be related to inertial instability in strongly anticyclonic flows (Arakawa 1952; Stone 1966; Knox 1997):

$$f \left[f \left(1 - \frac{1}{\text{Ri}} \right) + \zeta \right] < 0. \quad (A22)$$

j. Horizontal temperature gradient

The horizontal temperature gradient is a measure of the deformation and also vertical wind shear from the thermal wind relation, and is routinely used by airline forecasters. It was also used in Buldovskii et al. (1976):

$$|\nabla_H T| = \left[\left(\frac{\partial T}{\partial x} \right)^2 + \left(\frac{\partial T}{\partial y} \right)^2 \right]^{1/2}. \quad (A23)$$

k. Wind-related indices

Besides the speed vertical shear (A3), the wind speed

$$s = |\mathbf{v}| \quad (A24)$$

may be related to turbulence. Also Endlich (1964) noted that jet stream turbulence tended to be associated with both high wind speeds and large directional wind shears; thus,

$$s \left| \frac{\partial \psi}{\partial z} \right| \quad (A25)$$

could be a turbulence diagnostic, where ψ is the wind direction.

l. Dutton's empirical index

Dutton's empirical index (Dutton 1980) is based on linear regression analyses of turbulence reports over the North Atlantic and northwest Europe during 1976 and various synoptic-scale turbulence indices:

$$E = 1.25 S_H + 0.25 S_V^2 + 10.5, \quad (\text{A26})$$

where S_H is the horizontal wind shear,

$$S_H = \left(\frac{u}{s}\right) \frac{\partial s}{\partial y} - \left(\frac{v}{s}\right) \frac{\partial s}{\partial x}, \quad (\text{A27})$$

and the constants were empirically determined by Dutton.

m. MOS CAT probability predictor indices

The following indices were found to be the most skillful subset of a suite of turbulence diagnostics used in NCEP's Nested Grid Model (NGM) model output statistics (MOS) probability forecasts (Reap 1996):

$$|\mathbf{v}| \text{DEF} \quad (\text{A28})$$

and

$$\left| \frac{\partial T}{\partial z} \right| \text{DEF}. \quad (\text{A29})$$

n. Unbalanced flow

There is some evidence that regions of strong imbalance may be related to turbulence aloft (e.g., Knox 1997; McCann 2001; Koch and Caracena 2002). The UBF diagnostic (Knox 1997, 2001; McCann 2001; O'Sullivan and Dunkerton 1995; Koch and Caracena 2002) formulation used within GTG was developed by Koch and Caracena (2002) and McCann (2001), and derives from the residual of the nonlinear balance equation

$$\text{UBF} = -\nabla^2 \Phi + 2J(u, v) + f\zeta - \beta u, \quad (\text{A30})$$

where Φ is geopotential, J is the Jacobian operator, and β is the Coriolis frequency gradient.

The Lagrangian Rossby number RoL is another unbalanced flow diagnostic originally suggested by Van Tuyl and Young (1982). Following the definition of Ro in Bluestein (1992, Eq. 4.1.93) and evaluating in isentropic coordinates gives

$$\text{RoL} = \frac{|D\mathbf{v}/Dt|}{|f\mathbf{v}|} \approx \frac{|\mathbf{v} \cdot \nabla \mathbf{v}|}{|f\mathbf{v}|} \approx \frac{|-\nabla M - f\mathbf{k} \times \mathbf{v}|}{|f\mathbf{v}|}, \quad (\text{A31})$$

where M is the Montgomery streamfunction.

The hydrostatic NWP model derived or nonhydrostatic computed vertical velocity

$$|\mathbf{w}| \quad (\text{A32})$$

and the horizontal divergence,

$$\Delta_H = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}, \quad (\text{A33})$$

may also be measures of unbalanced flow.

Other unbalanced flow related diagnostics developed by McCann (2001) and used in a case study by Knox (2001) include

$$\text{ABSIA} = |v_i - v_c|^2, \quad (\text{A34})$$

where

$$v_i = |\mathbf{v} \cdot \nabla \mathbf{v}|/f \quad \text{and} \quad v_c = K_s |\mathbf{v}|^2/f,$$

and

$$\text{AGI} = \zeta_{\text{curv}} + f/2, \quad (\text{A35})$$

with

$$\zeta_{\text{curv}} = K_s |\mathbf{v}|,$$

and where K_s is the streamline curvature.

o. NCSU1

The NCSU1 is described in Kaplan et al. (2004) and was developed from investigations of several severe turbulence encounters:

$$\text{NCSU1} = \frac{1}{\text{MAX}(\text{Ri}, 10^{-5})} \text{MAX} \left(u \frac{\partial u}{\partial x} + v \frac{\partial v}{\partial y}, 0 \right) |\nabla \zeta|. \quad (\text{A36})$$

p. Negative vorticity advection

A rule-of-thumb forecasting approach used by the airlines is to look for regions of large negative vorticity advection (NVA) computed as follows (Bluestein 1992, p. 335):

$$\text{NVA} = \text{MAX} \left\{ \left[-u \frac{\partial}{\partial x} (\zeta + f) - v \frac{\partial}{\partial y} (\zeta + f) \right], 0 \right\}. \quad (\text{A37})$$

APPENDIX B

Thresholds Used with the 20-km RUC

Table B1 provides thresholds and default weights for the upper- and midlevel turbulence diagnostics used within the current version of GTG. The thresholds are determined from median values for 1 yr (2003) of

TABLE B1. Thresholds (T1, T2, T3, T4, and T5) corresponding to null, light, moderate, severe, and extreme turbulence categories (cf. Fig. 3) and default weights for the individual diagnostics currently used in GTG at upper and midlevels. The column labeled “Eq.” refers to the equation number in appendix A.

Index name	Eq.	Units	T1	T2	T3	T4	T5	Default wt
Upper levels								
–Ri	(A1)	—	–20	–2.0	–0.6	–0.3	0.5	0.103
CP	(A4)	kt ²	0.	1000	5000	12 000	30 000	0.095
EDR	(A7)	m ^{2/3} s ^{–1}	0.02	0.046	0.065	0.09	0.13	0.097
SIGW	(A8)	m s ^{–1}	0.02	0.035	0.065	0.10	0.60	0.097
DTF3	(A6)	m ² s ^{–2}	0.5	1.0	2.0	5.0	8.3	0.083
F _θ	(A9)	m ² s ^{–3} K ^{–2}	10 ^{–5}	4.5 × 10 ^{–5}	10 ^{–4}	3 × 10 ^{–4}	7.1 × 10 ^{–4}	0.121
TI1	(A15)	s ^{–2}	2.0 × 10 ^{–7}	10 ^{–6}	1.7 × 10 ^{–6}	3.0 × 10 ^{–6}	4.3 × 10 ^{–6}	0.109
∇ _H T	(A23)	K m ^{–1}	1.1 × 10 ^{–5}	2.6 × 10 ^{–5}	4.0 × 10 ^{–5}	6.5 × 10 ^{–5}	9.0 × 10 ^{–5}	0.111
UBF	(A30)	s ^{–2}	10 ^{–8}	1.8 × 10 ^{–8}	2.7 × 10 ^{–8}	4.0 × 10 ^{–8}	10 × 10 ^{–8}	0.088
NCSU1	(A36)	s ^{–3}	0.0	10 ^{–13}	3.5 × 10 ^{–13}	1.5 × 10 ^{–12}	4.0 × 10 ^{–12}	0.096
Midlevels								
EDR	(A7)	m ^{2/3} s ^{–1}	0.01	0.028	0.034	0.044	0.054	0.111
SIGW	(A8)	m s ^{–1}	0.02	0.061	0.11	0.18	0.25	0.102
F _p	(A10)	m ² s ^{–3} Pa ^{–2}	10 ^{–11}	4.510 ^{–11}	1.510 ^{–10}	5 × 10 ^{–10}	10 ^{–9}	0.116
TI1	(A15)	s ^{–2}	2.0 × 10 ^{–7}	5.7 × 10 ^{–7}	1.1 × 10 ^{–6}	2.7 × 10 ^{–6}	8.0 × 10 ^{–6}	0.110
∇ _H T	(A23)	K m ^{–1}	1.8 × 10 ^{–5}	3.5 × 10 ^{–5}	4.9 × 10 ^{–5}	7.0 × 10 ^{–5}	9.1 × 10 ^{–5}	0.109
v	(A24)	m s ^{–1}	9	18	23	29	35	0.107
v DEF	(A28)	m s ^{–2}	3.0 × 10 ^{–4}	1.1 × 10 ^{–3}	1.7 × 10 ^{–3}	2.5 × 10 ^{–3}	3.3 × 10 ^{–3}	0.126
ABSIA	(A34)	kt ² m ^{–1} s ^{–1}	3	7.9	15.2	34	60	0.110
NCSU1	(A36)	s ^{–3}	0.0	5.8 × 10 ^{–14}	10 ^{–12}	5 × 10 ^{–9}	10 ^{–7}	0.109

1800 UTC 6-h forecast (valid 0000 UTC) index–PIREP pairs for each of the five major turbulence categories. The default weights are derived from the area² under the PODY–PODN 6-h forecast curves (Figs. 5 and 6), normalized so that the resultant sum is unity.

REFERENCES

Arakawa, H., 1952: Severe turbulence resulting from excessive wind-shear in tropical cyclones. *J. Meteor.*, **9**, 221–223.

Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004: Mesoscale weather prediction with the RUC hybrid isentropic-terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473–494.

Bluestein, H. B., 1992: *Principles of Kinematics and Dynamics*. Vol. I. *Synoptic–Dynamic Meteorology in Midlatitudes*, Oxford University Press, 431 pp.

—, 1993: *Observations and Theory of Weather Systems*. Vol. II. *Synoptic–Dynamic Meteorology in Midlatitudes*, Oxford University Press, 594 pp.

Breiman, L., and J. Friedman, R. A. Olshen, and C. J. Stone, 1984: *Classification and Regression Trees*. Wadsworth International Group, 358 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, *Ninth Conf. on Aviation, Range, and Aerospace Meteorology*, Orlando, FL, Amer. Meteor. Soc., 393–398.

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

—, J. L. Mahoney, J. Henderson, T. L. Kane, R. Bullock, and J. E. Hart, 2000: The turbulence algorithm intercomparison

exercise: Statistical verification results. Preprints, *Ninth Conf. on Aviation, Range, and Aerospace Meteorology*, Orlando, FL, Amer. Meteor. Soc., 466–471.

Brown, R., 1973: New indices to locate clear-air turbulence. *Meteor. Mag.*, **102**, 347–360.

Buldovskii, G. S., S. A. Bortnikov, and M. V. Rubinshtejn, 1976: Forecasting zones of intense turbulence in the upper troposphere. *Meteor. Gidrol.*, **2**, 9–18.

Clark, T. L., J. R. Scoggins, and R. E. Cox, 1975: Distinguishing between CAT and non-CAT areas by use of discriminant functional analysis. *Mon. Wea. Rev.*, **103**, 514–520.

Colson, D., and H. A. Panofsky, 1965: An index of clear-air turbulence. *Quart. J. Roy. Meteor. Soc.*, **91**, 507–513.

Cornman, L. B., C. S. Morse, and G. Cuning, 1995: Real-time estimation of atmospheric turbulence severity from in-situ aircraft measurements. *J. Aircraft*, **32**, 171–177.

—, G. Meymaris, and M. Limber, 2004: An update on the FAA Aviation Weather Research Program’s *in situ* turbulence measurement and reporting system. Preprints, *11th Conf. on Aviation, Range, and Aerospace Meteorology*, Hyannis, MA, Amer. Meteor. Soc., CD-ROM, P4.3.

Dutton, J., and H. A. Panofsky, 1970: Clear air turbulence: A mystery may be unfolding. *Science*, **167**, 937–944.

Dutton, M. J. O., 1980: Probability forecasts of clear-air turbulence based on numerical output. *Meteor. Mag.*, **109**, 293–310.

Ellrod, G. P., and D. L. Knapp, 1992: An objective clear-air turbulence forecasting technique: Verification and operational use. *Wea. Forecasting*, **7**, 150–165.

Endlich, R. M., 1964: The mesoscale structure of some regions of clear-air turbulence. *J. Appl. Meteor.*, **3**, 261–276.

Fahey, T. H., 1993: Northwest Airlines atmospheric hazards advisory and avoidance system. Preprints, *Fifth Conf. on Aviation Weather Systems*, Vienna, VA, Amer. Meteor. Soc., 409–413.

- Frehlich, R., and R. Sharman, 2004a: Estimates of turbulence from numerical weather prediction model output with applications to turbulence diagnosis and data assimilation. *Mon. Wea. Rev.*, **132**, 2308–2324.
- , and —, 2004b: Estimates of upper level turbulence based on second order structure functions derived from numerical weather prediction model output. Preprints, *11th Conf. on Aviation, Range, and Aerospace Meteorology*, Hyannis, MA, Amer. Meteor. Soc., CD-ROM, P4.13.
- Hanley, J. A., and B. J. McNeil, 1982: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.
- Hopkins, R. H., 1977: Forecasting techniques of clear-air turbulence including that associated with mountain waves. WMO Tech. Note 155, 31 pp.
- Ihaka, R., and R. Gentleman, 1996: R: A language for data analysis and graphics. *J. Comput. Graph. Stat.*, **5**, 299–314.
- Jacobi, C., A. H. Siemer, and R. Roth, 1996: On wind shear at fronts and inversions. *Meteor. Atmos. Phys.*, **59**, 235–243.
- Kane, T. L., and B. G. Brown, 2000: Confidence intervals for some verification measures—A survey of several methods. Preprints, *15th Conf. on Probability and Statistics in the Atmospheric Sciences*, Asheville, NC, Amer. Meteor. Soc., 46–49.
- Kaplan, M. L., and Coauthors, 2004: Characterizing the severe turbulence environments associated with commercial aviation accidents. A real-time turbulence model (RTTM) designed for the operational prediction of hazardous aviation turbulence environments. NASA CR-2004-213025, 54 pp.
- , A. W. Huffman, K. M. Lux, J. J. Charney, A. J. Riordan, and Y.-L. Lin, 2005: Characterizing the severe turbulence environments associated with commercial aviation accidents. Part 1: A 44-case study synoptic observational analyses. *Meteor. Atmos. Phys.*, **88**, 129–153.
- Keller, J. L., 1990: Clear air turbulence as a response to meso- and synoptic-scale dynamic processes. *Mon. Wea. Rev.*, **118**, 2228–2242.
- Kelsch, M., C. Fischer, and J. L. Mahoney, 2004: Forecaster's evaluation of the integrated turbulence forecast algorithm (ITFA), Winter 2003. Preprints, *20th Conf. on Weather Analysis and Forecasting*, Seattle, WA, Amer. Meteor. Soc., CD-ROM, 7.1.
- Khari, V. V., and F. W. Zwiers, 2003: On the ROC score of probability forecasts. *J. Climate*, **16**, 4145–4150.
- Knox, J. A., 1997: Possible mechanism of clear-air turbulence in strongly anticyclonic flows. *Mon. Wea. Rev.*, **125**, 1251–1259.
- , 2001: The breakdown of balance in low potential vorticity regions: Evidence from a clear air turbulence outbreak. Preprints, *13th Conf. on Atmospheric and Oceanic Fluid Dynamics*, Breckenridge, CO, Amer. Meteor. Soc., 64–67.
- Koch, S. E., and F. Caracena, 2002: Predicting clear-air turbulence from diagnosis of unbalance flow. Preprints, *10th Conf. on Aviation, Range, and Aerospace Meteorology*, Portland, OR, Amer. Meteor. Soc., 359–363.
- Koshyk, J. N., and K. Hamilton, 2001: The horizontal energy spectrum and spectral budget simulated by a high-resolution troposphere–stratosphere–mesosphere GCM. *J. Atmos. Sci.*, **58**, 329–348.
- Kronebach, G. W., 1964: An automated procedure for forecasting clear-air turbulence. *J. Appl. Meteor.*, **3**, 119–125.
- Laikhtman, D. L., and Y. Z. Al'ter-Zalik, 1966: Use of aerological data for determination of aircraft buffeting in the free atmosphere. *Izv. Akad. Nauk SSSR. Fiz. Atmos. Okeana*, **2**, 534–536.
- Lane, T. P., J. D. Doyle, R. Plougonven, M. A. Shapiro, and R. D. Sharman, 2004: Observations and numerical simulations of inertia–gravity waves and shearing instabilities in the vicinity of a jet stream. *J. Atmos. Sci.*, **61**, 2692–2706.
- , R. Sharman, H.-M. Hsu, W. D. Hall, M. A. Shapiro, R. Plougonven, and J. J. Murray, 2005: Numerical simulations of gravity waves and turbulence during the ATReC campaign. *43rd AIAA Aerospace Science and Exhibit*, Reno, NV, American Institute of Aeronautics and Astronautics, AIAA-2005-262.
- Lee, Y.-G., B.-C. Choi, R. Sharman, G. Wiener, and H.-W. Lee, 2003: Determination of the primary diagnostics for the CAT (clear-air turbulence) forecast in Korea. *J. Kor. Meteor. Soc.*, **39**, 677–688.
- Leshkevich, T. V., 1988: Automated method of predicting the probability of clear-air turbulence. *Meteor. Gidrol.*, **10**, 27–33.
- Lester, P. F., 1994: *Turbulence: A New Perspective for Pilots*. Jeppesen Sanderson, 212 pp.
- Lindborg, E., 1999: Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence? *J. Fluid Mech.*, **388**, 259–288.
- Mahoney, J. L., J. K. Henderson, B. G. Brown, J. E. Hart, A. Loughie, C. Fischer, and B. Sigren, 2002: The Real-Time Verification System (RTVS) and its application to aviation weather forecasts. Preprints, *10th Conf. on Aviation, Range, and Aerospace Meteorology*, Portland, OR, Amer. Meteor. Soc., 323–326.
- Marroquin, A., 1998: An advanced algorithm to diagnose atmospheric turbulence using numerical model output. Preprints, *16th Conf. on Weather Analysis and Forecasting*, Phoenix, AZ, Amer. Meteor. Soc., 79–81.
- Marzban, C., 2004: The ROC curve and the area under it as performance measures. *Wea. Forecasting*, **19**, 1106–1114.
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- McCann, D. W., 2001: Gravity waves, unbalanced flow, and aircraft clear air turbulence. *Natl. Wea. Dig.*, **25**, 3–14.
- McCullagh, P., and J. A. Nelder, 1989: *Generalized Linear Models*. Chapman and Hall, 532 pp.
- O'Sullivan, D., and T. J. Dunkerton, 1995: Generation of inertia–gravity waves in a simulated life cycle of baroclinic instability. *J. Atmos. Sci.*, **52**, 3695–3716.
- Passetti, V., T. C. Carty, D. L. Simms, and J. A. Weinrich, 2000: Integrated Turbulence Forecasting Algorithm meteorological evaluation. Preprints, *Ninth Conf. on Aviation, Range, and Aerospace Meteorology*, Orlando, FL, Amer. Meteor. Soc., 472–475.
- Reap, R. M., 1996: Probability forecasts of clear-air turbulence for the contiguous U.S. National Weather Service Office of Meteorology Tech. Procedures Bull. 430, 15 pp.
- Ripley, B. D., 1996: *Pattern Recognition and Neural Networks*. Cambridge University Press, 415 pp.
- Roach, W. T., 1970: On the influence of synoptic development on the production of high level turbulence. *Quart. J. Roy. Meteor. Soc.*, **96**, 413–429.
- Schwartz, B., 1996: The quantitative use of PIREPs in developing aviation weather guidance products. *Wea. Forecasting*, **11**, 372–384.
- , S. G. Benjamin, S. M. Green, and M. R. Jardin, 2000: Ac-

- curacy of RUC-1 and RUC-2 wind and aircraft trajectory forecasts by comparison with ACARS observations. *Wea. Forecasting*, **15**, 313–326.
- Shapiro, M. A., 1978: Further evidence of the mesoscale and turbulence structure of upper level jet stream-frontal zone systems. *Mon. Wea. Rev.*, **106**, 1100–1111.
- Sharman, R., C. Tebaldi, and B. Brown, 1999: An integrated approach to clear-air turbulence forecasting. Preprints, *Eighth Conf. on Aviation, Range, and Aerospace Meteorology*, Dallas, TX, Amer. Meteor. Soc., 68–71.
- , —, J. Wolff, and G. Wiener, 2002: Results from the NCAR Integrated Turbulence Forecasting Algorithm (ITFA) for predicting upper-level clear-air turbulence. Preprints, *10th Conf. on Aviation, Range, and Aerospace Meteorology*, Portland, OR, Amer. Meteor. Soc., 351–354.
- Stone, P. H., 1966: On non-geostrophic baroclinic instability. *J. Atmos. Sci.*, **23**, 390–400.
- Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology*. Kluwer Academic, 670 pp.
- Tebaldi, C., D. Nychka, B. G. Brown, and R. Sharman, 2002: Flexible discriminant techniques for forecasting clear-air turbulence. *Environmetrics*, **13**, 859–878.
- Tung, K. K., and W. W. Orlando, 2003: The k -3 and k -5/3 energy spectrum of atmospheric turbulence: Quasigeostrophic two-level model simulation. *J. Atmos. Sci.*, **60**, 824–835.
- Van Tuyl, A. H., and J. A. Young, 1982: Numerical simulation of nonlinear jet streak adjustment. *Mon. Wea. Rev.*, **110**, 2038–2054.
- Vinnichenko, N. K., N. Z. Pinus, S. M. Shmeter, and G. N. Shur, 1980: *Turbulence in the Free Atmosphere*. Plenum, 310 pp.
- Vogel, G. N., and C. R. Sampson, 1996: Clear air turbulence indices derived from U.S. Navy numerical model data: A verification study. NRL/MR/7543-96-7223, Naval Research Laboratory, Monterey, CA, 30 pp.
- Weinrich, J. A., and D. Sims, 2002: Integrated turbulence forecasting algorithm 2001 meteorological evaluation. Preprints, *10th Conf. on Aviation, Range, and Aerospace Meteorology*, Portland, OR, Amer. Meteor. Soc., 299–302.