NEXRAD Turbulence Detection Algorithm (NTDA)
and
CIT Avoidance Guidelines and D-CIT Development

FY 2008 Year-End Status Report
on Turbulence RT Tasks 08.7.3.1.1, 08.7.3.1.2, and 08.7.3.13

Provided in fulfillment of
Deliverables 08.7.3.1.1E1, 08.7.3.1.2E1, and 08.7.3.13E1

Submitted by
The National Center for Atmospheric Research
Introduction

In FY 2008, NCAR was given Technical Direction by the FAA’s Aviation Weather Research Program Office to perform continued research and development and an operational demonstration of the NEXRAD Turbulence Detection Algorithm (NTDA) under task 08.7.3.1.1, “NTDA maintenance, verification and enhancement” and task 08.7.3.1.2, “NTDA Operational Demonstration”. NTDA research and development efforts focused on refactoring the NTDA software and redesigning the algorithm to accommodate changes to the NEXRAD data and improve algorithm performance. The FY08 operational demonstration followed successful NTDA demonstrations in FY05-FY07, but involved an expansion in the size of the EDR mosaic domain to nearly the entire CONUS to support DCIT and GTGN development. This year, the experimental cockpit uplinks performed in previous years were not funded. This report summarizes the research and development and operational demonstration activities in fulfillment of Turbulence Research Team (TRT) deliverables 08.7.3.1.1E1 and 08.7.3.1.2E1.

NCAR was also given Technical Direction to continue evaluation of the FAA thunderstorm avoidance guidelines and research and develop an automated algorithm (“DCIT”) to diagnose the likely location and intensity of convectively-induced turbulence (CIT). The directive for this effort was provided by Turbulence RT Task 08.7.3.13, “CIT avoidance guidelines and D-CIT development”. This report briefly summarizes the work performed in FY08 and also includes a conference paper describing both NTDA and DCIT status.
Summary of FY08 Efforts for Task 08.7.3.1.1: NTDA maintenance, verification and enhancement

Activities in FY08 included a significant re-factoring of the NTDA software and updating it to accommodate recent changes in the NEXRAD operational modes. Initially, work focused on finalizing the NTDA-2 refactoring, which was required to clean up code structure, increase maintainability and reduce memory and CPU usage. The NTDA-2 was then tested with numerous cases and compared against NTDA-1 to ensure the algorithm output was identical. Work then focused on updating the algorithm. The NEXRAD Open Radar Products Generator (ORPG) Build 10 uses a new message 31 data type in place of the previous message 1, so the NTDA software was updated to accommodate both message 1 and message 31 data, allowing full backwards compatibility for retrospective case studies. Additional changes were made for super-resolution data, dual-pol fields (present in message 31 but not yet populated with real data from operational NEXRADs), clutter contamination and improved quality control. Once clutter information was properly decoded, a number clutter cases were analyzed to determine the best use of these data for mitigating spectrum width contamination. Currently, quality control procedures are being separated into those that identify contamination that increases spectrum width bias (e.g., insect returns, clutter or overlaid echoes) from those that affect its variance (primarily SNR). Because NEXRADs employ numerous operational modes that have substantial differences in their spectrum width statistics, simulation studies have been performed to allow the quality control procedures to be customized for each such operational mode. Refinement and implementation of this updated design are continuing.
Summary of FY08 Efforts for Task 08.7.3.1.2: NTDA operational demonstration

The primary activity under this task has been to expand the real-time NEXRAD data ingest and processing to cover the entire CONUS, with a total of 133 NEXRADs now being employed. Sample output is shown in Figure 1. It became apparent after adding the additional NEXRADs that the data processing load was too great for our initial setup and number of servers. An additional server was purchased and added to the system. Also, enhancements were made to the mosaic that allowed multiple sub-domain mosaics to run independently with a final CONUS mosaic created from stitching the sub-domains together. These changes have lead to a stable real-time, operational demonstration system. An added feature of this year’s demonstration is the computation of “echo tops” fields based on the reflectivity mosaic for 10 dBZ, 18 dBZ and 30 dBZ thresholds, along with “turbulence tops” fields for light, moderate and severe turbulence (Figures 2-4). These fields provide useful summaries of the 3-D mosaic information both for human analysis and for incorporation into DCIT.

Analysis of poor performance cases has lead to the discovery numerous issues with individual radars providing bad or unexpected ORPG Level II data. Some of these were fixed by informing the ROC and others were dealt with by adding specific quality control checks. The NTDA reflectivity and turbulence 3-D mosaic data grids were again supplied to users via a web-based display. Direct data feeds, including both the 3-D mosaics and the derived echo tops and turbulence tops fields, continue to supply data for use in DCIT and GTGN development. Real-time NTDA data from 34 NEXRADs are also being provided to the National Severe Storm Laboratory to support development and testing of a prototype operational 3-D EDR mosaic in conjunction with Advanced Weather Radar Techniques RT Tasks 08.6.29 and 08.6.34.
Figure 1: NTDA operational demonstration research display images. (Top) Reflectivity mosaic at 2145 UTC 07/27/2008, altitude 33,000 feet; (Bottom) NTDA EDR mosaic at the same time and altitude.
Figure 2: NTDA operational demonstration research display images for the same case as Figure 1. (Top) Reflectivity echo tops computed using a 10 dBZ threshold; (Bottom) “Light turbulence tops” field computed from NTDA EDR mosaic with an intensity threshold of 0.10 m$^{1/3}$ s$^{-1}$. Note that there is generally good correspondence between reflectivity of 10 dBZ or greater and light or greater turbulence for this case.
Figure 3: NTDA operational demonstration research display images for the same case as Figure 1.  (Top) Reflectivity echo tops computed using an 18 dBZ threshold; (Bottom) “Moderate turbulence tops” field computed from NTDA EDR mosaic with an intensity threshold of 0.30 m$^{3/3}$ s$^{-1}$.  Note that moderate turbulence is often found in a subset of regions with reflectivity of 18 dBZ or greater, but also occasionally in regions with lower reflectivity. Comparison with Figure 4 also shows that regions with reflectivity 30 dBZ or greater do not consistently contain regions of detected moderate-or-greater turbulence.
Figure 4: NTDA operational demonstration research display images for the same case as Figure 1. (Top) Reflectivity echo tops computed using a 30 dBZ threshold; (Bottom) “Severe turbulence tops” field computed from NTDA EDR mosaic with an intensity threshold of 0.50 m$^{\frac{1}{3}}$ s$^{-1}$. 
Summary of FY08 Efforts for Task 08.7.3.13: CIT avoidance guidelines and D-CIT development

FY08 activities for this task continued to focus on development of a MySQL database that allows United Airlines in situ eddy dissipation rate (EDR) reports to be compared with fields and features derived from radar, satellite, lighting, and Rapid Update Cycle (RUC) model data and derived diagnostics. The initial effort has involved compiling a comprehensive set of data fields for three months in the summer of 2007; after these data are analyzed to determine what fields are most relevant, the database will be extended to summer 2008 and then to winter months.

A random forest analysis was used to create an empirical model for predicting CIT based on a subset of fields and features. Real-time data feeds were created to supply necessary radar, model and satellite fields to the algorithm. This DCIT prototype algorithm has been running reliably in real-time since early July, producing deterministic and probabilistic assessments of CIT at 1 km vertical intervals with the same horizontal resolution as the RUC-13 grid. Tuning these deterministic and probabilistic assessment fields is currently underway. Custom software was written to run the random forest in parallel on multiple threads, thereby exploiting the multi-core architecture of modern servers and reducing the processing time considerably. This software has enabled the random forest empirical model to run with an update frequency of 15 minutes. An image showing DCIT output is included in the attached conference paper.
Conference Paper

The conference paper attached to this report provides additional status information for the NTDA and DCIT tasks. The citation for this paper is:

Remote Detection and Diagnosis of Thunderstorm Turbulence

John K. Williams*, Robert Sharman, Jason Craig, and Gary Blackburn
National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307

ABSTRACT

This paper describes how operational radar, satellite and lightning data may be used in conjunction with numerical weather model data to provide remote detection and diagnosis of atmospheric turbulence in and around thunderstorms. In-cloud turbulence is measured with the NEXRAD Turbulence Detection Algorithm (NTDA) using extensively quality-controlled, ground-based Doppler radar data. A real-time demonstration of the NTDA includes generation of a 3-D turbulence mosaic covering the CONUS east of the Rocky Mountains, a web-based display, and experimental uplinks of turbulence maps to en-route commercial aircraft. Near-cloud turbulence is inferred from thunderstorm morphology, intensity, growth rate and environment data provided by (1) satellite radiance measurements, rates of change, winds, and other derived features, (2) lightning strike measurements, (3) radar reflectivity measurements and (4) weather model data. These are combined via a machine learning technique trained using a database of in situ turbulence measurements from commercial aircraft to create a predictive model. This new capability is being developed under FAA and NASA funding to enhance current U.S. and international turbulence decision support systems, allowing rapid-update, high-resolution, comprehensive assessments of atmospheric turbulence hazards for use by pilots, dispatchers, and air traffic controllers. It will also contribute to the comprehensive 4-D weather information database for NextGen.

Keywords: Atmospheric turbulence, aviation safety, NextGen, thunderstorms, convection, convectively-induced turbulence (CIT), Doppler weather radar, NEXRAD

1. INTRODUCTION

Studies have shown that turbulence in and around thunderstorms may be responsible for over 60% of turbulence-related aircraft accidents[1][2]. According to Federal Aviation Administration (FAA) guidelines, aircraft must circumnavigate thunderstorms by wide margins both horizontally and vertically to mitigate the risk of encountering the dangerous turbulence the storms may generate. In practice, interpretation of these guidelines is subjective and limited by available weather information, and the guidelines may make large regions of airspace unavailable to aircraft on days of widespread convection. In order to provide pilots, dispatchers and air traffic managers a more precise assessment of the turbulence location and severity—and maintain safety while minimizing unnecessary disruptions to air traffic—the FAA and NASA have sponsored research aimed at detecting turbulence inside thunderstorms and diagnosing turbulence in the near-storm environment, called convectively-induced turbulence (CIT). The NEXRAD Turbulence Detection Algorithm (NTDA) makes use of data from the U.S. network of operational Doppler weather radars, performing extensive quality control and producing estimates of eddy dissipation rate (EDR), an atmospheric turbulence metric. NTDA provides information on turbulence intensity in regions where there is sufficient cloud droplet density or precipitation to obtain good signal-to-noise ratio (SNR) measurements, and where contamination from overlaid echoes or non-atmospheric returns is minimal. Thus, NTDA may detect in-cloud turbulence, but turbulence outside the cloud boundary is not directly measured and must be inferred by other means. An automated algorithm for diagnosis of convectively-induced turbulence, called DCIT, is being developed to combine information from both observations and numerical weather prediction (NWP) model assessments of environmental conditions to produce high-resolution, rapid-update, 3D probabilistic assessments of light, moderate, and severe turbulence. Input data used by DCIT include fields and derived feature variables from NEXRAD reflectivity mosaics, Geostationary Operational Environmental Satellite (GOES) radiances, U.S. National Lightning Detection Network (NLDN) data, and Rapid Update Cycle (RUC) model forecasts. DCIT is scheduled to be incorporated into the comprehensive Graphical Turbulence Guidance Nowcast (GTGN) for dissemination via the National Weather Service (NWS) Aviation Weather Center’s Aviation Digital Data Service (ADDS), and may also provide input to the FAA’s Consolidated Storm Prediction for Aviation (CoSPA) product[3].

* jkwillia@ucar.edu; phone 303-497-2822; fax 303-497-8401; www.ucar.edu
Ultimately, DCIT will contribute to the Joint Planning and Development Office’s (JPDO’s) vision of a comprehensive weather information database for all aviation users to support the next-generation air transportation system (NextGen), providing valuable strategic and tactical decision support to pilots, dispatchers and air traffic controllers.

This paper describes the NTDA algorithm and a real-time CONUS demonstration of the NTDA including a web-based display and cockpit uplinks to select en-route United Airlines aircraft. It also explains how an empirical modeling (a.k.a. data mining, statistical analysis, or machine learning) technique has been used to develop empirical models that associate near-cloud CIT with fields and derived features from thunderstorm observations and RUC model fields. The detection and diagnosis algorithms’ skill is evaluated based on statistical comparisons with in-situ EDR reports from commercial aircraft, and sample output for several case studies is shown.

2. NEXRAD TURBULENCE DETECTION ALGORITHM (NTDA)

2.1 Overview

The NTDA has been developed at NCAR during the past several years under direction and funding by the FAA’s Aviation Weather Research Program (AWRP) with the goal of using the nation’s network of operational Doppler weather radars—called Weather Surveillance Radar 88 Doppler (WSR-88D) or Next-Generation Radar (NEXRAD)—to directly detect turbulence in clouds and thunderstorms that may be hazardous to aviation. In February 2007, the NTDA software was delivered to the National Weather Service Radar Operations Center for inclusion in the NEXRAD Open Radar Product Generator (ORPG), and it is being deployed in the spring and summer of 2008 as part of the new NEXRAD software baseline. The NTDA is a fuzzy-logic algorithm that uses radar reflectivity, radial velocity, and spectrum width data to perform data quality control and compute eddy dissipation rate (EDR), an aircraft-independent atmospheric turbulence metric, along with an associated confidence (EDC). For locations where there is a sufficiently strong radar return (e.g., in clouds and precipitation) and the spectrum width contamination is not too large, the EDC values are close to one and the data may be used with high confidence. Both EDR and EDC are produced for each elevation tilt on a polar grid with 1 degree azimuth and 2 km range spacing. Once NTDA data transmission and distribution is established, the NTDA data from each NEXRAD could be made available to any interested users, in addition to the FAA AWRP turbulence products.

For the past three years, a real-time demonstration of the NTDA has been run at NCAR over increasingly larger domains: after beginning with 16 radars in the summer of 2005, it now utilizes 133 NEXRADs that cover nearly the entire CONUS. The raw data from the radars are ingested, processed using the NTDA software, and then merged to form a 3-D mosaic. Cockpit uplink messages showing in-cloud turbulence ahead are generated for all United Airlines aircraft in the demonstration domain, and may be uplinked to pilots who have registered their flights on an NCAR website. The NTDA has been statistically verified using comparisons to collocated in situ turbulence reports, and subjectively validated based on pilot feedback.

2.2 NTDA Algorithm

The NTDA algorithm has been described in some detail in earlier work. An updated description of its major elements is provided below. The input to the NTDA is the raw “Level II” data, which include the reflectivity (DZ), radial velocity (VE) and spectrum width (SW). These fields are related to the 0th, 1st and 2nd moments of the Doppler spectrum, respectively.

1. The raw radar data are censored to remove measurements that appear so contaminated that they are judged to have no value to the algorithm. These include occasional “ring” artifacts of spurious values and “sun spikes” generated when the radar points near to the direction of the sun.

2. A set of “interest maps” are used to assess the quality of each SW measurement based on its estimated signal-to-noise ratio (SNR), overlaid power ratio (PR), the local spatial variance of the SW field, the deviation of the SW measurement from a local linear fit of its neighbors, the associated DZ and height above ground (to reduce spurious returns from insects in and near the boundary layer), the likelihood that the measurement is contaminated by uncompensated Anomalous Propagation (AP) clutter returns as identified by the Radar Echo Classifier, whether clutter filtering was applied to the measurement, whether the pixel overlays one from another trip in which clutter filtering has been applied (in which case the PR estimated from the DZ field is known to be in error), the radial range of the measurement from the radar, and the range of the measurement relative to the unambiguous range (the largest range at which radar echoes may be received before the next transmission). These interest values are combined via a weighted
geometric average to produce an overall 0 to 1 “confidence” value for each SW measurement. The interest map details are dependent on the Volume Coverage Pattern (VCP) being employed by the NEXRAD, since each VCP uses a different operational mode (e.g., dwell time per measurement) and hence has different error statistics for SW. The interest maps used by the NTDA have been determined based on both simulation results and empirical data.

(3) The SW measurements are squared and scaled into $\varepsilon^{2/3}$ (the 2/3 power of eddy dissipation rate) by multiplying by a range-dependent factor with a form determined by turbulence theory\(^6\) and modified based on empirical verification data. A local confidence-weighted average is taken over a 2-km radius disc around each grid point, followed by a square root, to generate the output EDR field (really the cubed root of eddy dissipation rate, $\varepsilon^{1/3}$, which appears to be approximately proportional to human assessments of aircraft turbulence encounter severity). The associated confidence field, EDC, is computed via a local average of the associated SW confidences.

A number of additional enhancements will be required in the near future to keep the NTDA functioning well given the many planned changes to the NEXRAD hardware and Open Radar Data Acquisition (ORDA) system, which will include the addition of dual-polarization capability and new operational modes.

2.3 NTDA Real-time Processing

The NTDA demonstration system uses data from the 133 NEXRADs depicted in Figure 1, which cover nearly the entire CONUS. The raw Level II data is obtained in real-time from Integrated Radar Data Services (IRaDS) and ingested using Unidata’s Local Data Manager (LDM) software. Each elevation tilt is processed separately, with output produced on a polar grid. Every five minutes, the latest data from each radar in the demonstration domain are merged onto a mosaic grid having 15 vertical levels at intervals of 3,000 ft starting at 3,000 ft. The horizontal resolution of the grid is 0.02° latitude by 0.02° longitude, or approximately 2 km $\times$ 2 km. The mosaicking algorithm works by computing the latitude, longitude and mean sea-level (MSL) altitude of each spherical-coordinate radar data point using a standard beam bending model to account for the refractive index gradient. Each measurement having a confidence above some threshold is then incorporated into a distance- and confidence-weighted average at each nearby mosaic grid point. The distance weighting function used is determined by the VCP being employed, and is designed to interpolate vertically and use a range-dependent Gaussian-shaped smoothing kernel in the horizontal. Additionally, the height of the maximum radar measurement for each column in the grid is computed. This height field is then smoothed and used as an estimate to the detectable cloud top height. Grid points above the estimated cloud top height are assigned a confidence of zero and are flagged as “bad” data. This last step is useful in mitigating the radar “ring” artifacts that otherwise often appear in the mosaic, particularly at the upper altitudes. In addition, “tops” products are computed for regions of EDR > 0.1 m$^{2/3}$ s$^{-1}$ (roughly, light or greater turbulence), EDR > 0.3 m$^{2/3}$ s$^{-1}$ (moderate or greater turbulence), and EDR > 0.5 m$^{2/3}$ s$^{-1}$ (severe turbulence). These 2D maps, similar to NEXRAD echo tops products, provide useful summary information regarding the potential in-cloud turbulence hazard.

Figure 1: The 133 NEXRADs in the 2008 NTDA demonstration (blue diamonds) and those not used (yellow). Of those not used, only the one in Maine and the one in southern Kentucky have real-time Level II data available.
2.4 NTDA Demonstration

The 3D NTDA demonstration mosaic data is made available to airline dispatchers, Central Weather Service Unit forecasters, and other interested users via a password-protected, web-accessible Java display similar to the visualization tool in Experimental ADDS\cite{7} (http://www.weather.aero). This tool overlays the NTDA EDR, confidence, or reflectivity mosaic grids over a U.S. map, and also provides access to several Experimental ADDS products including Graphical Turbulence Guidance (GTG)\cite{8} turbulence forecasts, along with RUC model temperature, humidity, and wind speed. Wind barbs and PIREPs may optionally be overlaid, as can real-time in situ turbulence reports obtained via the FAA’s automated EDR reporting system.\cite{9} For purposes of display, the EDR is scaled into turbulence severity categories associated with NTDA EDRs falling in the light or moderate turbulence categories. Perhaps more significantly, regions where radar reflectivity is sufficiently strong and where spectrum width measurement contamination is minimal, it is typical for much of the display to be shaded grey, representing “No Data,” as in the example displayed in Figure 2. It should be emphasized that “No Data” does not mean “no turbulence”; rather, no information at all is provided by the NTDA in these regions. It is notable that the regions of reflectivity above 20 dBZ or so are often—but not always—associated with NTDA EDRs falling in the light or moderate turbulence categories. Perhaps more significantly, regions of quite low reflectivity are sometimes associated with moderate or even severe turbulence, such as over central Wisconsin in this case. This observation underscores the fact that, while reflectivity is a good proxy for thunderstorm intensity, it is not precisely correlated with the associated convective turbulence hazard. The NTDA’s contribution to characterizing the environmental situation is that it measures the turbulence itself that can be a significant hazard to safe flight.

Figure 2: (Upper left) The NTDA demonstration system Java display showing a plan view of the EDR mosaic at 24,000 ft altitude, with United Airlines flight tracks depicting in situ EDR measurements overlaid: blue dots represent null reports, while green and orange squares denote turbulence encounters. (Lower left) The corresponding reflectivity mosaic plan view. (Right) An annotated cockpit uplink printout depicting regions in-cloud turbulence ahead of a United Airlines flight. The flight information at top has been deliberately obscured.
Following the creation of each 3D EDR mosaic, the FAA’s Aircraft Situation Display to Industry (ASDI) data are used to identify the position and planned route for every United Airlines aircraft within the uplink domain, and a customized text-based message is generated to depict the NTDA turbulence detected along the next 100 nautical miles of its flight path, 40 nmi to either side, and 9,000 ft. above and below. If the pilot has registered the flight through an NCAR website, the message is evaluated to determine whether it meets turbulence severity and coverage criteria for uplink. If it does, the message is uplinked to the cockpit via an ARINC data link, and prints out on the cockpit’s Aircraft Communications Addressing and Reporting System (ACARS) printer. An example message is shown on the right side of Figure 2. Following the flight, the pilot may log on to the NCAR website to review the text messages generated for the flight, including those that did not merit uplinking, and provide feedback on the timeliness, accuracy and utility of those that were uplinked. United pilots participating in the cockpit uplink demonstration have reported that the NTDA uplinks accurately depict the in-cloud turbulence the aircraft is flying through, and in particular are much better correlated with turbulence intensity than the reflectivity shown on their airborne radar display.\textsuperscript{[10]} The cockpit uplink demonstration indicates that the NTDA data can be provided to users in time to be tactically useful, increasing pilot’s situational awareness of a potentially hazardous situation.

3. DIAGNOSING CONVECTIVELY-INDUCED TURBULENCE (CIT)

3.1 Overview
While the NTDA uses Doppler weather radar data to detect in-cloud turbulence hazards, it does not give any information in regions with low-SNR radar returns, contaminated data or poor radar coverage. On the other hand, the GTG product mentioned earlier makes use of NWP model data, but due to limitations of latency, resolution, and model physics, does not explicitly address turbulence associated with thunderstorms. For these reasons, an FAA and NASA-funded effort is underway to use thunderstorm features derived from various observations and NWP model data representing the storm environment to infer locations in and around storms where turbulence is likely to exist. The new CIT diagnosis product resulting from this research, called DCIT, will be incorporated into a new “nowcast” version of GTG called GTG Nowcast (GTGN). Proper diagnosis of this important turbulence source will potentially improve airline safety and may also help mitigate the significant delays that now frequently afflict the national airspace system during periods of widespread convection.

Thunderstorms can induce major disturbances to their surroundings, including changes in stability, winds and windshear, but the precise mechanisms for the generation and propagation of CIT are not currently well-understood. Clearly, the horizontal shears created by localized updrafts and downdrafts are one source of turbulence. Fine-scale numerical modeling studies have suggested that turbulence—particularly, above clouds—may be caused by gravity waves generated by convection that then propagate above and away from the storm before breaking. For instance, Lane et al.\textsuperscript{[11]} performed simulations of a commercial flight accident case in which severe turbulence was encountered above isolated deep convection. The study suggested that gravity waves were produced above the rapidly growing cloud as it penetrated the lower stratosphere; they then propagated several km into the stratosphere and “broke” to produce turbulence in the clear air above the cloud. This mechanism is dependent on environmental conditions, which could either promote or inhibit the propagation and breaking of the gravity waves\textsuperscript{[12]}. Other studies have shown that convective outflow from either isolated thunderstorm cells or large mesoscale convective systems can modify the static stability and vertical shear within the upper troposphere, lowering the Richardson number and producing turbulence in certain locations relative to the convection. For all of these sources, it seems likely that the intensity of CIT is related to the size, depth, intensity and longevity of a thunderstorm as well as features of the near-storm environment including stability, strength of environmental upper-tropospheric vertical shear and the interaction of the thunderstorm outflow with ambient winds. Numerical modeling case studies are still far from systematically or quantitatively characterizing these relationships, though they do suggest important features that are likely relevant. Thus, empirical models are sought to associate thunderstorm observations and RUC model data with \textit{in situ} reports of turbulence intensity to provide the required diagnostic capability.

3.2 Data sources
The advent of automated, quantitative turbulence reports from commercial aircraft has been essentially a prerequisite to developing empirical diagnostic models for CIT. The pilot report (PIREP) data used in the creation of GTG’s clear-air turbulence diagnostics are subjective and, more importantly, have time and position errors that render them virtually useless for characterizing the typically highly-localized, transient CIT.\textsuperscript{[13]} In contrast, the FAA’s automated \textit{in situ}
Several sources of data are available for providing thunderstorm characteristics and environmental state variables that may be expected to be related to the incidence of CIT. One source of information about the location and severity of thunderstorms is the National Convective Weather Detection (NCWD) product\[15\]. Briefly, the NCWD provides a 2-D CONUS mosaic of convective intensity in units of vertically integrated liquid (VIL, kg m$^{-2}$) on a 4-km grid. The NCWD makes use of NEXRAD VIL data along with cloud-to-ground lightning data from the National Lightning Detection Network (NLDN)\[16\]. VIL data at locations having radar echo tops below 15,000 ft. are removed. The number of lightning strikes over the past 10 minutes occurring within 8 km of a grid point are then combined with the latest VIL mosaic using an empirically-derived formula to create the NCWD convective intensity value. This NCWD grid is updated every 5 minutes. In addition to VIL, a Unisys 2-D mosaic of NEXRAD echo tops data was available, as were GOES IR satellite data and a lightning density field derived from the NLDN data.

Near-storm environment data were provided by the Rapid Update Cycle (RUC) NWP model\[17\]. The RUC data include 13-km 2-D and 3-D grids of variables including winds, turbulent kinetic energy (TKE), convective available potential energy (CAPE), convective inhibition (CIN), potential temperature, humidity mixing ratio, and a number of others. Additionally, all of the RUC-derived turbulence diagnostics used in the GTG forecast algorithm, which was referenced earlier, were computed. These include Richardson number (Ri), structure function eddy dissipation rate (EDR), horizontal and vertical shear, inverse stability, and a large number of others.

To produce training and testing data sets, five months of data from June - October 2005 were used. Each in situ EDR measurement was associated with collocated feature variables containing environment and thunderstorm information. RUC and RUC-derived data from the closest model analysis time were interpolated from the nearest points surrounding the aircraft location. The GOES IR temperature nearest the point was also used, as was the radar echo top data for the nearest mosaic grid point. The distance to the nearest NCWD VIL value above each of several selected convective intensity thresholds (0.14, 0.76, 3.5, 6.9, 12 and 32 kg m$^{-2}$) were also computed. These various feature variables do not represent independent predictors of turbulence; rather, they are quite highly correlated with one another. The prediction of turbulence must be achieved by an empirical function of the joint distribution of the variables, which can be built using a machine learning data fusion method as described below.

### 3.3 Random forests

A statistical analysis and machine learning data fusion technique called “random forests”\[18\] has been used to analyze the feature variables described in the previous section and develop the required empirical turbulence diagnosis model. Essentially, a random forest is a group of decision trees trained in such a fashion that they are weak and weakly-correlated and hence may function as an ensemble of experts by “voting” on the correct classification of a given input. The use of an ensemble of such trees creates an empirical model that minimizes the risk of overfitting the training set, a significant problem with individual decision trees that requires techniques such as “pruning” to obtain good generalization performance. In constructing each tree of a random forest, one begins with a “training set” containing many instances of the candidate feature variables along with an associated “truth” value (here the in situ peak EDR measurement). While it is being trained to create an empirical model, the random forest also produces an estimate of the classification importance of each variable, which can provide a helpful starting point for comparing the potential contributions of different feature variables and selecting a minimal, skillful set.

Once a random forest has been trained, the trees function as an “ensemble of experts” to make predictions. For example, Figure 1 shows a conceptual diagram of a random forest with 100 trees. When a new data instance (or “data point” or “feature vector”: a set of feature variable values at the point for which the diagnosis is being made) is presented, each tree will perform a classification (here the EDR bin number). These classification “votes” are then compiled, and can be used to classify the point based on the consensus “winner”, or the vote distribution may be used to derive a probability
for each possible class. For example, if 40 trees vote “0” (no hazardous turbulence) and 60 vote “1” (turbulence), one might be able to scale the 60% classification confidence for hazardous turbulence into a probability estimate.

**Figure 3:** Conceptual diagram of a random forest, an ensemble of weak, weakly-correlated decision trees that “vote” on the classification of each new instance (data point).

### 3.4 RF training results: the value of observation data for DCIT

In order to evaluate the potential contribution of observation-derived data to the model-based forecasts provided by GTG, the random forest was trained with model-based feature variables alone and with both model and observation data. Only *in situ* EDR reports above 15,000 ft and within 40 nmi of convection (NCWD VIL > 3.5 kg m⁻²) were included in this analysis in order to focus on CIT prediction instead of a mixture of CIT and CAT. The number of votes for turbulence categories above light, moderate, and severe turbulence thresholds (chosen as EDR values of 0.1, 0.3, and 0.5 m²/³ s⁻¹) were then analyzed to produce receiver operating characteristic (ROC) curves [19]. The ROC curves are produced by choosing in turn all possible fractions of votes for turbulence above the specified level and, using the *in situ* EDR reports as “truth”, determining the probability of a proper diagnosis of turbulence (PoDY) and the probability of properly diagnosing the absence of turbulence (PoDN). The PoDY values were then plotted against the associated values of the quantity 1 – PoDN, also called the probability of false detection, to form the ROC curves shown in Figure 4. Because the curves on the right-hand plot come much closer to the ideal point, (0,1), the random forest trained using both observation and model data has clearly superior performance. Note that the ROC curve for severe-or-greater turbulence is the best in both plots, which is encouraging since severe turbulence is most operationally significant. It is also the most “jagged” because severe turbulence measurements are relatively quite rare in the test data set.

**Figure 4:** ROC curves showing the performance of a random forest empirical model for diagnosing summertime turbulence within 40 nmi of convection when trained with only model-based feature variables (left) or with both the model and observation-based features described in the text (right).
3.5 DCIT training results: feature field importance

Next, the variable importance results produced by the random forest trained using both model and observation based features was analyzed. The results are shown in Table 1. The top two feature variables in the random forest importance list are the GOES IR minus flight level temperature, which is related to the aircraft’s vertical position relative to the cloud top, and the GOES IR value itself, which is related to the thunderstorm top’s altitude and thus its intensity. The features ranked 3, 4, 6, 9 and 11 are RUC-based turbulence diagnostics that were originally designed to diagnose clear-air turbulence (CAT) but may also include some CIT when the thunderstorm influence is captured by the RUC model forecast. The other top fields are related to the aircraft’s horizontal distance to thunderstorm pixels of various intensities. The features ranked 22, 23 and 24 are related to the aircraft’s altitude, which in turn is climatologically related to the likelihood of turbulence. Number 25 again reflects the altitude of the aircraft relative to the cloud top, but the NEXRAD reflectivity-based echo top field, which is quantized at 5,000 ft. increments, evidently contributes much less than the GOES IR field to the empirical model’s predictive skill.

Table 1: Importance ranks and values for 16 of the top 25 feature fields for diagnosing CIT.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Impt.</th>
<th>Diagnostic Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.88</td>
<td>GOES IR minus RUC-derived Flight Level Temperature</td>
</tr>
<tr>
<td>2</td>
<td>3.43</td>
<td>GOES IR</td>
</tr>
<tr>
<td>3</td>
<td>2.80</td>
<td>RUC-derived Structure Function EDR</td>
</tr>
<tr>
<td>4</td>
<td>2.60</td>
<td>RUC-derived Ellrod Index</td>
</tr>
<tr>
<td>5</td>
<td>1.83</td>
<td>Distance to VIL &gt; 3.5 kg m⁻²</td>
</tr>
<tr>
<td>6</td>
<td>1.77</td>
<td>RUC-derived Saturated Richardson Number</td>
</tr>
<tr>
<td>7</td>
<td>1.61</td>
<td>Distance to VIL &gt; 6.9 kg m⁻²</td>
</tr>
<tr>
<td>8</td>
<td>1.57</td>
<td>Distance to VIL &gt; 0.9 kg m⁻²</td>
</tr>
<tr>
<td>9</td>
<td>1.55</td>
<td>RUC-derived Frontogenesis Function</td>
</tr>
<tr>
<td>10</td>
<td>1.54</td>
<td>Distance to VIL &gt; 12.0 kg m⁻²</td>
</tr>
<tr>
<td>11</td>
<td>1.47</td>
<td>RUC-derived Vertical Shear</td>
</tr>
<tr>
<td>12</td>
<td>1.42</td>
<td>Distance to VIL &gt; 30.0 kg m⁻²</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>0.78</td>
<td>RUC MSL Pressure minus RUC Flight Level Pressure</td>
</tr>
<tr>
<td>23</td>
<td>0.76</td>
<td>RUC Pressure</td>
</tr>
<tr>
<td>24</td>
<td>0.74</td>
<td>RUC-derived Temperature</td>
</tr>
<tr>
<td>25</td>
<td>0.73</td>
<td>NEXRAD Echo Top minus Geopotential Height</td>
</tr>
</tbody>
</table>

3.6 DCIT real-time prototype

A random forest empirical model trained to diagnose turbulence based on NWP model fields and derived variables, along with observed thunderstorm features, is currently running in real-time at NCAR’s Research Applications Laboratory. This prototype DCIT algorithm produces both a deterministic EDR estimate as well as probability estimates for light-or-greater, moderate-or-greater and severe-or-greater turbulence derived from the random forest votes. An example screen shot from the research display depicting the probability of moderate-or-greater turbulence is shown in Figure 5. It is clear that many of the flights are navigating around the areas where the probability of turbulence is assessed as high. However, in the case of the flight from eastern Texas into southeastern New Mexico, it appears from the DCIT assessment that the large deviation may be unwarranted given the relatively small threat. Further data collection and verification will be necessary to determine whether the DCIT diagnoses are reliable, and to refine and improve the empirical predictive model.
Figure 5: Sample output from a prototype random forest-based DCIT algorithm running in real-time at NCAR/RAL, depicting the estimated probability of encountering moderate-or-greater turbulence at 37,000 ft. Overlaid are United Airlines B-757 flight tracks for aircraft flying between 35,000 and 40,000 ft, some of which are clearly deviating around the regions identified by DCIT as potentially hazardous.

4. SUMMARY AND CONCLUSIONS

This paper has described how operational Doppler weather radar data may be used to directly detect in-cloud turbulence via the NEXRAD Turbulence Detection Algorithm, and how a real-time mosaic based on this capability was made available to en-route commercial airline pilots and shown to have good accuracy and utility. However, the NTDA is unable to detect convectively-induced turbulence outside of clouds, or in regions where NEXRAD coverage is poor or the data are contaminated. To fill this significant gap, thunderstorm observations and environmental features are used to attempt to infer where turbulence is likely to exist near storms. A machine learning approach was used to combine observation data from operational Doppler radar, GOES satellites and the National Lightning Detection Network in conjunction with numerical weather model data to develop an empirical predictive model for CIT. The combination of these two systems will provide a valuable enhancement to tactical thunderstorm turbulence assessments, thereby contributing an important new capability for the FAA’s Graphical Turbulence Guidance Nowcast product and, eventually, the comprehensive 4-D weather information database planned for NextGen.

ACKNOWLEDGEMENTS

This research was funded in part by a NASA Cooperative Agreement Notice award from the Science Mission Directorate’s Applied Sciences Program.

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 or position of the FAA.

In addition to the authors, several other scientists and engineers have contributed important elements of this work. These include Jennifer Abernethy, Steve Carson, Larry Cormman, Andy Cotter, Danika Gilbert, Lisa Goodrich, Seth Linden, Dr. Greg Meymaris, Dr. Stan Trier, and Jaimi Yee at NCAR, Wayne Feltz, Kris Bedka and their co-workers at the
University of Wisconsin-Madison, Dr. John Mecikalski and his colleagues at the University of Alabama-Huntsville, and Dr. Rob Fovell at UCLA.

REFERENCES