1.2

A HYBRID MACHINE LEARNING AND FUZZY LOGIC APPROACH TO CIT DIAGNOSTIC DEVELOPMENT

John K. Williams*, Jason Craig, Andrew Cotter, and Jamie K. Wolff
National Center for Atmospheric Research, Boulder, Colorado

1. INTRODUCTION

Convectively-induced turbulence (CIT) is one of several threats that requires aircraft to avoid thunderstorms in order to mitigate the risk of passenger injury or aircraft damage. Current Federal Aviation Administration (FAA) thunderstorm avoidance guidelines proscribe flight within 20 nautical miles of a thunderstorm, above thunderstorm tops or beneath anvils. 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. An automated turbulence product that makes use of radar, lightning, satellite, numerical weather model and convective nowcast data to objectively diagnose the likelihood of turbulence in the near-storm environment could provide valuable strategic and tactical decision support to pilots, dispatchers and air traffic controllers.

The advent of automated, quantitative turbulence reports from commercial aircraft has made it possible to use machine learning techniques to help develop such diagnostics. This paper describes the use of a machine learning method called random forests—ensembles of weakly-correlated decision trees—to help establish relationships between storm features and aircraft turbulence that may then be used to develop a fuzzy logic predictive algorithm for turbulence intensity near thunderstorms. Values from the Rapid Update Cycle (RUC) numerical weather prediction model were interpolated to the aircraft position, and a spatial "dartboard" oriented relative to the wind direction was used to collect data on storm intensity and coverage in the vicinity of the aircraft. Random forests were trained to learn a predictive algorithm based on these quantities. In the process, they provided lists of the variables that were most useful in distinguishing different categories of turbulence, allowing subsequent simplification of the feature set without degrading predictive performance. In future work, the behavior of the random forest will be analyzed and the results used to develop a fuzzy logic algorithm that predicts turbulence based on thunderstorm features and environmental conditions. This fuzzy logic algorithm should be sufficiently efficient to run in a real-time system.

* *Corresponding author address:* John K. Williams, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307; email: jkwillia@ucar.edu.

2. THE CIT AVOIDANCE PROBLEM

Current FAA Aviation Weather Research Program turbulence research and development is focused in three areas: forecasting turbulence based primarily on Numerical Weather Prediction (NWP) model data, establishing routine and accurate reports of en route turbulence encountered by aircraft, and detecting turbulence remotely using Doppler radars and other sensors (Sharman et al. 2006a). Recently, an effort has begun to synthesize these approaches in order to generate a rapid-update turbulence "nowcast" product based on the latest NWP model and available *in situ* and remote sensor data. An important goal of this nowcast will be to provide short-term predictions of convectively-induced turbulence (CIT)—turbulence in and around thunderstorms—which studies have shown to be responsible for over 60% of turbulence-related aircraft accidents (Cornman and Carmichael 1993; see also Kaplan et al. 2005). Accurate diagnosis of this important source of turbulence will improve airline safety and also help mitigate the significant delays that now frequently afflict the national airspace system during periods of widespread convection.

The mechanisms for the generation and propagation of CIT are not currently well-understood by researchers. As the FAA thunderstorm avoidance guidelines indicate, CIT is commonly thought to be related to the proximity (vertical and horizontal), intensity, depth and extent of convection. The guidelines include the following:

- *Don't attempt to fly under a thunderstorm even if you can see through to the other side. Turbulence and wind shear under the storm could be disastrous.*
- *Do avoid by at least 20 miles any thunderstorm identified as severe or giving an intense radar echo. This is especially true under the anvil of a large cumulonimbus.*
- *Do clear the top of a known or suspected severe thunderstorm by at least 1,000 feet altitude for each 10 knots of wind speed at the cloud top.*
- *Do circumnavigate the entire area if the area has 6/10 thunderstorm coverage.*
- *Do regard as extremely hazardous any thunderstorm with tops 35,000 feet or higher whether the top is visually sighted or determined by radar.*

(source: FAA Advisory Circular 00-24, available at www.airweb.faa.gov/Regulatory_and_Guidance_Library /rgAdvisoryCircular.nsf/, and FAA Aeronautical Information Manual section 7-1-30, available from www.faa.gov/atpubs/AIM/). Numerical simulation studies have suggested that at least some CIT may be caused by gravity waves, including those generated by overshooting thunderstorm tops. For instance, Lane et al. (2003) performed simulations for an 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 propagated several km into the stratosphere and "broke" to produce turbulence in the clear air above the cloud. The incidence of CIT is also thought to be related to environmental conditions, which could either promote or inhibit the propagation and breaking of gravity waves, for instance. Nevertheless, understanding of this phenomenon is still quite tentative, and while the NEXRAD Turbulence Detection Algorithm (NTDA, Williams et al. 2005) will soon provide routine detection of turbulence in thunderstorm clouds, no operational system yet exists that reliably predicts out-of-cloud CIT.

3. FUZZY LOGIC AND RANDOM FORESTS

Fuzzy logic provides a principled way to encode expert knowledge or heuristics into algorithms that mimic a human's approach to solving challenging problems by weighing different sources of information and making judgments based on a preponderance of the evidence. The fuzzy logic approach is frequently used to develop decision support systems for weather hazards, including turbulence. For example, the Graphical Turbulence Guidance (GTG) system (Sharman et al. 2006b) combines a number of turbulence diagnostics based on NWP model data, with weights that are dynamically tuned based on their recent performance as measured against pilot reports (PIREPs). More generally, fuzzy logic algorithms may be constructed by surveying human experts on their approach to recognizing relevant patterns in data and using them to produce diagnoses or forecasts. The various parameters of these fuzzy logic algorithms may then be tuned using empirical data, if they are available, to optimize the algorithm's performance.

Unfortunately, the CIT diagnosis problem does not appear amenable to this approach because the phenomenology of CIT is so poorly understood. Without an expert available to identify the most significant variables and define at least the basic structure of an algorithm, it is not immediately clear how a fuzzy logic algorithm for CIT might be constructed except by naïve trial-and-error. In order to resolve this conundrum, the present paper proposes a hybrid algorithm development approach: using a machine learning algorithm to train an "expert" model based on available data, then utilizing the "learned" mapping to help inform the development of a fuzzy logic algorithm that can be employed in practice. The machine learning algorithm should identify the most important or useful feature variables for predicting the aircraft-measured turbulence so that a simple yet effective fuzzy logic algorithm can be developed. The model it provides for predicting turbulence can also subsequently be interrogated to determine the relationships of the various feature variables to the likely intensity of turbulence. If all goes well, this process should uncover patterns and relationships that might not be easily discerned using other data analysis methods, thereby helping researchers better understand the mechanisms underlying CIT in addition to providing a practical warning system.

The machine learning technique selected for the present study was random forests (Breiman 2001). Essentially, random forests are ensembles of weak, weakly-correlated decision trees that "vote" on the correct classification of a given input. These ensembles minimize the risk of overfitting the training set, a significant and well-known problem with individual decision trees. In constructing each tree of a random forest, a "bagged" training sample is selected by drawing a random subset of *n* elements from the *n*-member training set, with replacement after each draw. Then, at each node of the tree, a subset of *m* feature variables are selected as candidates for splitting, and the best of the candidates is used; this contrasts with the usual practice in decision tree construction in which the best of all feature variables is selected for splitting. Because not all feature variables are used to train each tree, those not used (the so-called "out-of-bag" samples) may be used to evaluate the performance of that tree. Leaving out feature variables in turn by setting them to their median value, their importance may be determined based on the degradation in classification performance of the subset of the random forest for which they were not used in training. Using this technique, the feature variables may be ranked in order of their importance to the random forest's performance.

4. DATA SOURCES

The turbulence "truth" data used for training and evaluation of the random forest consists of quantitative measurements from the FAA's automated *in situ* turbulence reporting system (Cornman et al. 1995 and 2004), currently operational on United Airlines B-737 and B-757 aircraft. The system provides reports of eddy dissipation rate (EDR, $\varepsilon^{1/3}$), an aircraft-independent atmospheric turbulence metric, at approximately one-minute intervals, including both the median and 90th percentile ("peak") EDR encountered over that period. The EDR data are reported in bins at 0.05 (roughly, null turbulence), 0.15 and 0.25 (light turbulence), 0.35 and 0.45 (moderate turbulence), 0.55 and 0.65 (severe turbulence), and 0.75 $m^{2/3}$ s^{-1} (extreme turbulence). The present study utilizes the peak EDR value because it supplies a good indication of hazard to the aircraft and is better distributed over the reporting bins. The location of the aircraft EDR report is taken to be the midpoint of each 1-minute flight segment; since commercial aircraft typically fly at airspeeds near 250 m s^{-1}, the peak EDR locations may be in error by 4 nmi (7.4 km) or more. Nevertheless, these uncertainties are significantly less that those of pilot reports (PIREPs) which frequently involve significant errors in the reported event's location and time (Schwartz 1996). The high temporal and spatial resolution and objective nature of the *in situ* EDR reports make them ideal for the present study.

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. Information about the location and severity of thunderstorms was derived from radar and lightning data via the National Convective Weather Detection (NCWD) product (Megenhardt et al. 2004). Briefly, the NCWD provides a 2-D 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 via a 2-D mosaic supplied by Unisys along with cloud-to-ground lightning data from the National Lightning Detection Network (NLDN). 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 combined with the latest VIL mosaic using an empirically-derived formula to create the NCWD convective intensity value. The 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

Environmental state data were provided by the Rapid Update Cycle (RUC) NWP model (Benjamin et al. 2004). 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 (Sharman et al. 2006b) 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 sets, five months of data from June – October 2005 were used. Each *in situ* EDR measurement was associated with feature variables containing environment and thunderstorm information. RUC and RUC-derived data from the nearest 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 intensity thresholds were computed, as were the echo tops associated with those VIL values. More challenging was associating the aircraft measurement with information about the coverage and relative proximity of convection relative to the prevailing wind direction. This was accomplished by orienting a polar coordinate system relative to the aircraft-measured wind vector (in an operational system, RUC or radar-derived winds would be used). Figure 1 depicts six 60° "wedges" oriented relative to the wind vector. To supply information on the proximity and relative location of convection, the closest distance to NCWD VIL above each of several intensity thresholds was computed for each of these wedges. Percent coverage data were then compiled by computing the proportion of VIL values above several selected intensity thresholds falling within prescribed range intervals in each wedge, i.e., the regions defined by the "dartboard" pattern defined in Figure 2. Cotter et al. (2007) describes a computationally efficient method for computing these values. All together, 272 feature variables were available to the machine learning algorithm, including the aircraft altitude, pressure corrected altitude, satellite IR data at the aircraft position, 24 RUC variables, 28 RUC-derived variables, 197 VIL distance and dartboard coverage values, and 20 echo top variables.

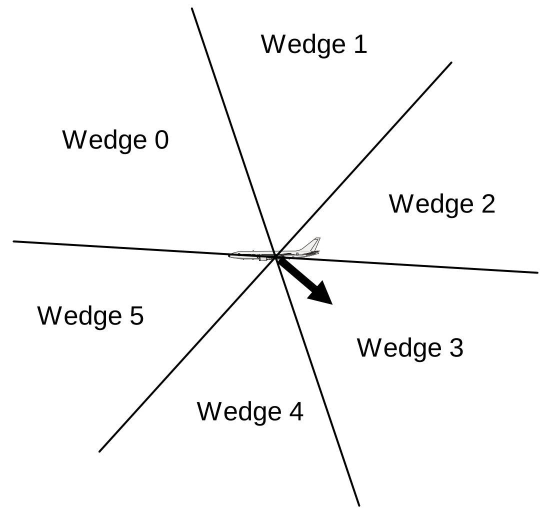

Figure 1: "Wedges" oriented relative to the wind vector (arrow) for purposes of collecting thunderstorm proximity information.

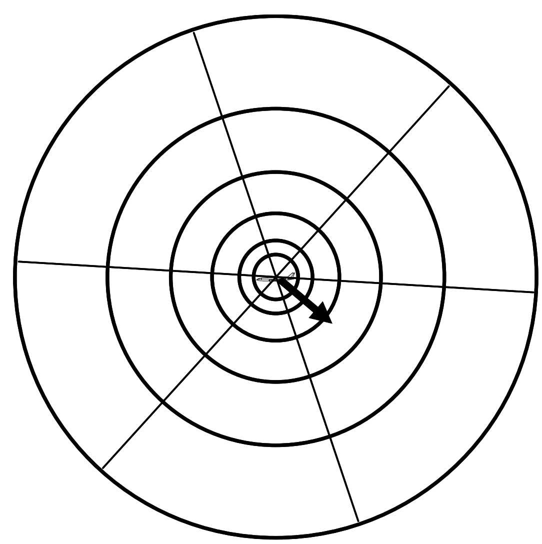

Figure 2: "Dartboard" oriented relative to the wind vector (arrow) for purposes of collecting thunderstorm intensity and coverage information. Range rings are at distances of 5, 10, 20, 40, 80 and 160 nmi from the aircraft.

Before being used for training, this dataset was analyzed to confirm that the feature variables were indeed related to the frequency and intensity of CIT. Figure 3 depicts the distribution of distances to convection having NCWD VIL > 3.5 kg m^{-2} for each of several levels of measured peak EDR. The increasing risk of turbulence encounters as the aircraft nears the thunderstorm is clearly evident. Figure 4 depicts the distribution of turbulence encounters as a function of the aircraft's distance above the NEXRAD echo top; negative values represent encounters below the echo top. Again, there is a clear relationship between this vertical proximity and the risk of turbulence. Finally, Figure 5 and Figure 6 depict histograms of dartboard coverage by NCWD VIL above various intensity thresholds for the regions enclosed by the six wedges and the 10 and 20 nmi range rings (see Figure 2). Figure 5 shows histograms for cases in which the *in situ* EDR is 0.05 (null turbulence), and Figure 6 shows histograms for EDRs of 0.35 or greater (moderate or greater turbulence). The difference in the shapes of these conditional histograms clearly indicates that the VIL coverages of these regions are positively correlated with the occurrence with turbulence. Similar conditional histograms for selected RUC and RUC-derived fields may be found in Sharman et al. (2006b).

The clear relationships between the measured variables and incidence of turbulence illustrated by these histograms suggest that they should be useful in predicting the location and intensity of turbulence. However, these variables do not represent *independent* predictors of turbulence; rather, they are quite highly correlated with one another. Simple averaging or summing of predictions from individual variables is therefore unlikely to produce an optimal turbulence diagnosis algorithm. Instead, the optimal prediction is a function of the joint distribution of the variables; for instance, some feature variables may have predictive value only within certain domains of others. Fortunately, learning such a function is precisely the domain in which a random forest can be effective. Decision trees effectively split up the hypercube of feature variable values and predict the class for each sub-region. Taking an ensemble of such trees exploits this feature while also making it possible to generalize well to an independent testing dataset.

5. RANDOM FOREST TRAINING AND RESULTS

The first step in running the random forest was to create appropriate training and testing sets. In order to minimize unwanted dependence between the variables used, separate days were chosen for the training and testing sets. Roughly one of every three days was chosen for training, and two for testing, each with even distributions over the period for which data were available. Because the "null" and "light" turbulence categories had by far the highest frequency in the dataset, the training set sampled only a limited number from these categories while utilizing all of the moderate or greater reports. The next step was to determine how many variables should be randomly selected as candidates for splitting at each node, and how many trees to grow. After some experimentation, the number of variables for splitting was chosen to be 25. Figure 7 shows the maximum true skill score achieved by the random forest as a function of the number of trees trained for one of the scenarios described below. It can be seen that the random forest's performance is quite near its maximum after just 100 trees, after which it continues to improve gradually as more trees are added. The results presented in this paper are based on training 500 trees.

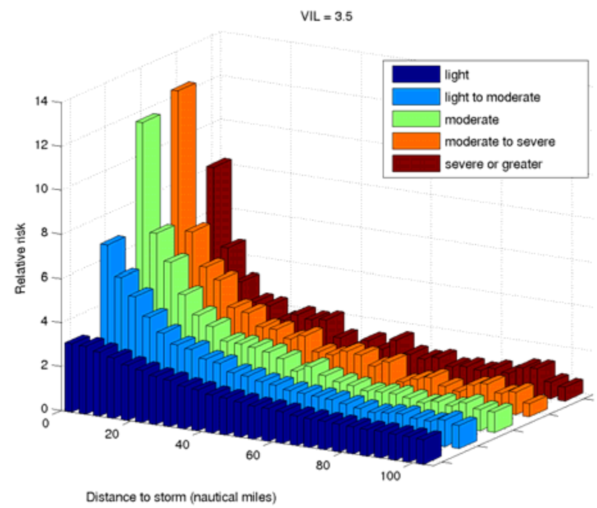

Figure 3: Distribution of distances to convection having NCWD VIL > 3.5 kg m^{-2} for various levels of peak EDR (dark blue = 0.15, light blue = 0.25, green = 0.35, orange = 0.45, and dark red = 0.55 or greater). The *z*-axis is normalized to show "relative risk", that is, the frequency of that level of turbulence divided by its overall frequency in the dataset.

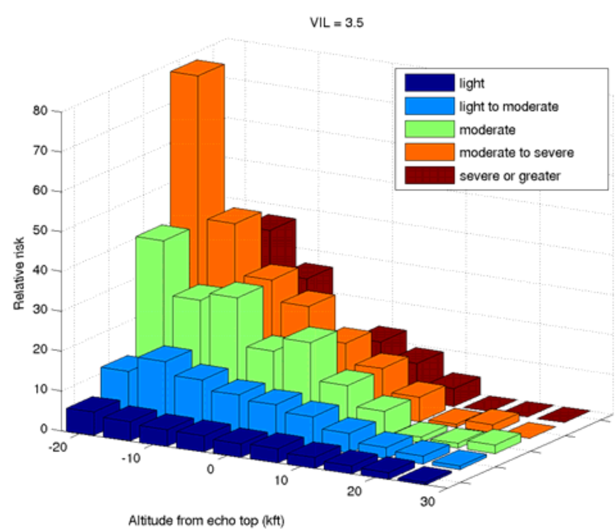

Figure 4: Distribution of vertical proximities to radar echo tops for various levels of turbulence intensity. The *z*-axis is again in terms of "relative risk."

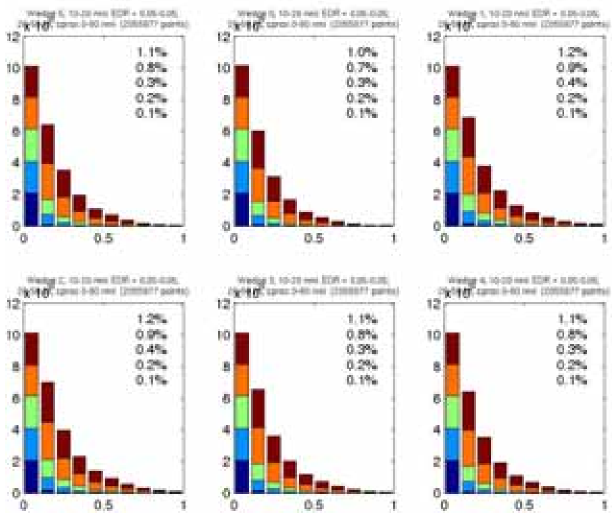

Figure 5: Histograms of dartboard wedge-range region coverage of NCWD VIL values above various thresholds (red = 0.9, orange = 5, green = 10, light blue = 15 and dark blue = 30 kg m^{-2}) for cases in which the *in situ* peak EDR was reported as 0.05 (null turbulence). Beginning with the upper middle plot and proceeding clockwise, the histograms are for regions between 10 and 20 nmi in wedges 0, 1, 2, 3, 4 and 5, as described in Figure 1. The percentages displayed on each histogram represent the average coverage of VIL above 0.9, 5, 10, 15 and 30 kg m^{-2}, respectively, in that region. Finally, values for the first bin of each histogram represent 1/100 of the actual frequency.

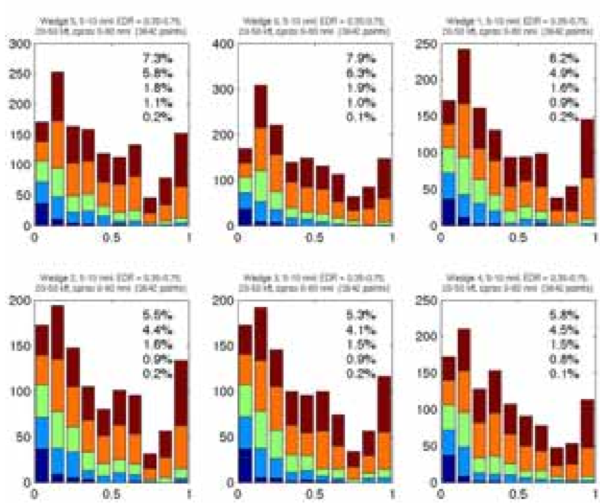

Figure 6: Conditional histograms like those in Figure 5, except for cases in which the *in situ* peak EDR was reported as 0.35 or greater, representing moderate or greater turbulence. The difference in the conditional histograms indicates that convective "coverage" might be a good predictor of turbulence.

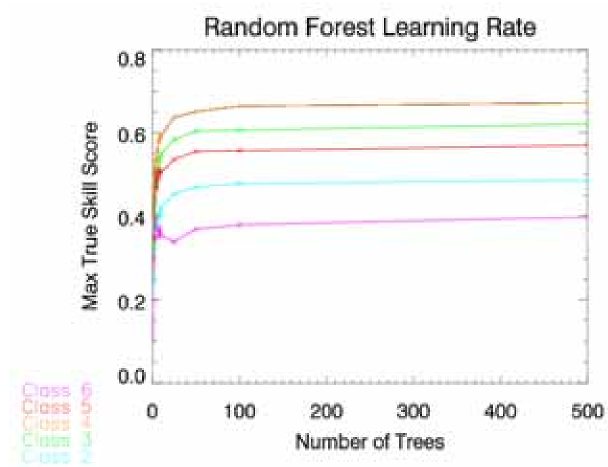

Figure 7: Plot showing maximum true skill score (TSS) as a function of the number of trees trained, with results for each of the various EDR categories.

Random forests were trained for a number of different scenarios and evaluated by generating ROC curves for different turbulence intensity levels like the ones in (figure). These curves were generated from the random forest output by taking the weighted mean of the turbulence category votes as a scalar predictor. A threshold on this predictor was selected, and the number of points correctly and incorrectly classified were used to determine the probability of detection (POD, *y*-axis) and probability of false detection (POFD, *x*-axis). As the threshold was varied, these pairs traced out the curves shown on the ROC plots. A POD of 1 and POFD of 0 would represent a perfect algorithm, so the curves that approach the upper-left corner most closely are deemed to show the best performance. The maximum true skill score (TSS = POD – POFD) was also computed as a summary statistic for the random forest's performance at each turbulence level.

Initial experiments with the random forest utilized EDR reports from all altitudes, resulting in quite exceptional ROC curves like the one shown in Figure 8 for training and testing on data from July 2005. However, further analysis of the data suggested that this performance was largely due to the vastly increased incidence of light or greater turbulence at levels below 15,000 ft, which made the aircraft altitude a powerful predictor. Furthermore, the *in situ* EDR reports are believed to be frequently inaccurate at these lower levels (Larry Cornman, personal communication), so it is not clear that how meaningful these results are. Finally, turbulence above 15,000 ft is of greatest significance to commercial airlines since passengers are less likely to be buckled in. For these reasons, further experiments with random forests focused on upper levels only. Unfortunately, the NEXRAD echo top data described earlier were inadvertently omitted from the training and testing datasets and so were not used for the results shown below.

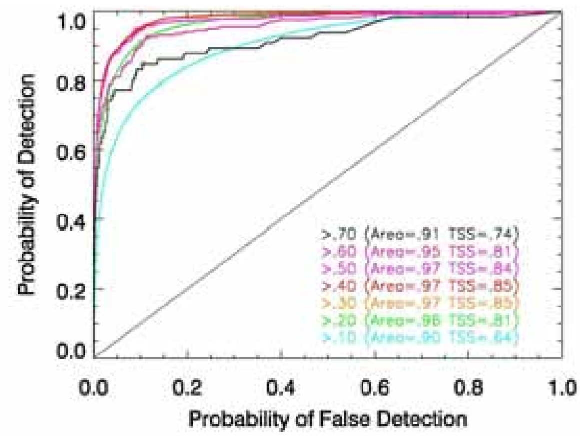

Figure 8: ROC curves showing results for a random forest trained and tested using data from July 2005. ROC curves are color-coded based on the EDR turbulence level being predicted. The area under the curve and the maximum TSS for each turbulence level are included in the legend.

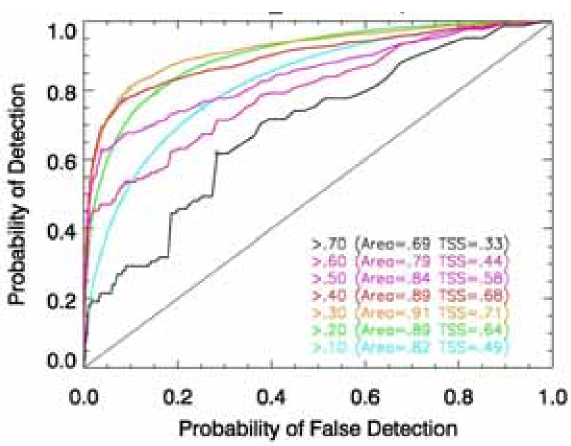

Figure 9: Same as Figure 8 for a random forest using data from June-October 2005, including only aircraft turbulence reports above 15,000 ft.

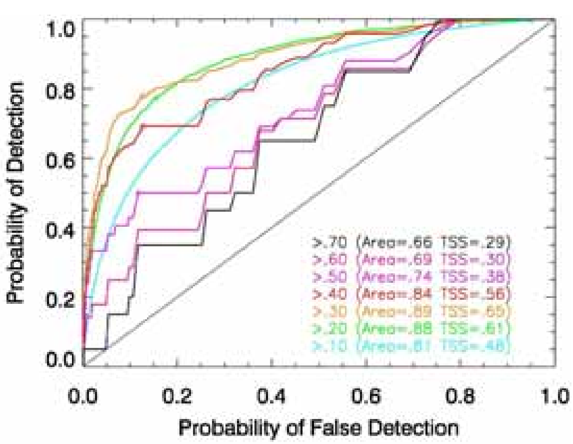

Figure 10: Same as Figure 9 but for a random forest trained using only the aircraft altitude, RUC, and RUC-derived feature variables and none of the radar-derived proximity, intensity and coverage variables.

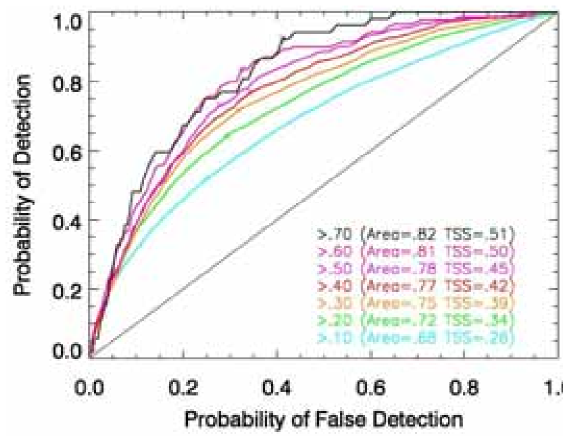

Figure 11: Same as Figure 9, but using only EDR reports within 40 nmi of convection as defined by NCWD VIL > 0.9 kg m^{-2}.

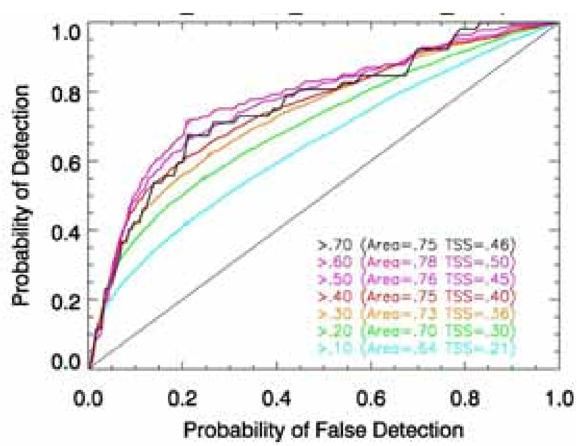

Figure 12: Same as Figure 10, but using only EDR reports within 40 nmi of convection as defined by NCWD VIL > 0.9 kg m^{-2}.

Figure 9 shows performance results for a random forest trained using data from June-October 2005, including only aircraft turbulence reports above 15,000 feet. The performance of the random forest as a predictive algorithm, evaluated on an independent "testing" subset of the data, is excellent, particularly for distinguishing the operationally significant moderate-or-greater (MoG, >0.30) turbulence from less than moderate turbulence. For MoG turbulence, a POD of 80% may be attained with a corresponding POFD of less than 10%, or a POD of 90% with POFD just over 20%. Figure 10 depicts the performance of a random forest trained without the NCWD VIL proximity, intensity and coverage data; that is, it utilized only data available to GTG. The performance is slightly worse, indicating that the NCWD did provide some useful information, though surprisingly little. Finally, Figure 11 shows results for a random forest trained and tested only on EDRs recorded within 40 nmi of convection, as defined by NCWD VIL > 0.9 kg m^{-2}, and Figure 12 shows results for a random forest trained and tested on EDRs within

40 nmi of convection but without the NCWD VIL features. The poorer performance of the random forests for the subset of cases within 40 nmi of thunderstorms is likely due to the fact that, since proximity is a good predictor of turbulence, the discrimination problem is much harder in this domain. In other words, the frequency of null turbulence reports is much smaller near thunderstorms, and discriminating different levels of non-null turbulence from one another is much more difficult. The random forest's performance is degraded somewhat when the NCWD data is omitted, but again not nearly as much as might have been expected.

As mentioned earlier, a random forest produces a ranked list of the feature variables that it finds most important to classifying the data during training. This list may be quite helpful in designing a fuzzy logic turbulence prediction algorithm because it indicates what fields should be used and which may be safely ignored in the interest of simplifying the algorithm. For the random forest whose results are displayed in Figure 9, 29 of the top 30 fields were RUC or RUC-derived variables, followed by a mix of additional RUC fields and distances to different NCWD convection intensities along the different wedges. Even for the data within 40 nmi of convection, none of the highest-ranked fields were related to NCWD intensity, coverage or proximity. These results are surprising, because, as Figure 3 shows, thunderstorm proximity is clearly related to the frequency of elevated turbulence reports, and one would expect that the 4-km NCWD, produced at 5-minute intervals, would be better at locating thunderstorms that the 13-km RUC analysis data, produced only hourly. Nevertheless, it is possible that the RUC data is capturing both the features most relevant to clear air turbulence as well as some of the conditions most related to CIT, rendering the NCWD data less important. If so, this is an important result, as it suggests that the effort required to incorporate NCWD data into a CIT warning system may not be worthwhile, at least for predicting upper-level turbulence. However, it is expected that the NCWD will be an important predictor at lower altitudes, and that the NEXRAD echo top data, when included among the feature variables, may significantly increase the random forest's performance.

6. SUMMARY AND FUTURE WORK

The causes of convectively-induced turbulence (CIT) are currently poorly understood, but it is a significant hazard to aviation that a turbulence nowcast system should be able to predict in order to be effective. In the absence of human expert knowledge to form the basis of designing a fuzzy logic algorithm to predict CIT, a machine learning technique may be used to develop a model that can then inform the development of a fuzzy logic algorithm by identifying important feature variables. By using a random forest instead of more common techniques such as linear correlation analysis, complex relationships between the feature variables and the incidence of turbulence may also be uncovered. Random forests are more straightforward to apply than other machine learning techniques, requiring significantly less expertise and "art" than artificial neural networks, for instance. Only two parameters really need to be established for a random forest: the number of candidate features for splitting at each node and the total number of trees to be trained. Furthermore, unlike individual decision trees, random forests do not tend to overfit the training set. Finally, the random forest provides a ranking of the feature variables in the order of importance that may be used as a starting place for developing an alternative algorithm.

In the present paper, we have applied the random forest to environmental information and turbulence diagnostics derived from a NWP model analysis; thunderstorm proximity, intensity and coverage data obtained from the NCWD grid, satellite IR data, NEXRAD echo tops. The performance of the random forests were quite impressive, but they showed, surprisingly, that the NCWD data did not appear to add significant information to the RUC data for predict the measured *in situ* aircraft turbulence, at least for upper levels. Unfortunately, the completeness of these results was limited by the inadvertent omission of radar echo tops data, which may be expected to be quite important in diagnosing CIT. Nevertheless, they point out the importance of using a machine learning algorithm to investigate whether data provide information useful to the solution of a problem before embarking on creating a solution.

The next phase of the project, after the missing echo tops data are added and the analyses above repeated, will be to run the trained random forest for a number of case studies. By running the random forest on each point in a grid, spatial patterns of predicted turbulence relative to thunderstorms will become evident. The proximity, extent and intensity of the thunderstorm will then be systematically varied, along with the RUC's environmental variables, in an attempt to uncover the significant relationships learned by the random forest model. From these, a fuzzy logic algorithm will be formulated. Instead of computing proximity, intensity and coverage features surrounding each grid point, the fuzzy logic algorithm might identify thunderstorm "objects" or environmental regions conducive to CIT and predict turbulence in and around them based on appropriate heuristics derived from the relationships learned by the random forest. The parameters of this algorithm could then be tuned using empirical data.

Note*: The latest version of this paper may be obtained from www.rap.ucar.edu/staff/williams/papers/ or by contacting the first author.*

7. ACKNOWLEDGEMENTS

The authors wish to express our appreciation to the Turbulence PDT's en-route turbulence project, including Larry Cornman, Martha Limber, and Sue Dettling, for supplying the *in situ* EDR data used in this study. And many thanks to Dan Megenhardt and the Convective Weather PDT for providing the NCWD product.

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.

8. REFERENCES

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

Breiman, L., 2001: Random forests. *Machine Learning*, **45**, 5-32.

Cornman, L. B. and B. Carmichael, 1993: Varied research efforts are under way to find means of avoiding air turbulence. *ICAO Journal*, **48**, 10-15.

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

Cornman, L. B., G. Meymaris and M. Limber, 2004: An update on the FAA Aviation Weather Research Program's *in situ* turbulence measurement and reporting system. *11th AMS Conference on Aviation, Range, and Aerospace Meteorology*.

Cotter, A., J. K. Williams, R. K. Goodrich and J. Craig, 2007: A Random Forest Turbulence Prediction Algorithm. *5th AMS Conference on Artificial Intelligence Applications to Environmental Science*, 1.3.

Kaplan, M. L., 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 analysis. *Meteor. Atmos. Phys.*, **88**, 129-153.

Lane, T. P., R. D. Sharman, T. L. Clark, and H.-M. Hsu, 2003: An investigation of turbulence generation mechanisms above deep convection. *J. Atmos. Sci.*, **60**, 1297-1321.

Megenhardt, D. L., C. Mueller, S. Trier, D. Ahijevych, and N. Rehak, 2004: NCWF-2 Probabilistic Forecasts. *11th AMS Conference on Aviation, Range, and Aerospace Meteorology*, 5.2.

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

Sharman, R. D., L. Cornman, J. K. Williams, S. E. Koch, and W. R. Moninger, 2006a: The AWRP Turbulence PDT. *12th AMS Conference on Aviation, Range, and Aerospace Meteorology*, 3.3.

Sharman, R., C. Tebaldi, G. Wiener and J. Wolff, 2006b: An integrated approach to mid-and upper-level turbulence forecasting. *Weather and Forecasting*, **21**, 268-287.