2.4 DEVELOPING A GLOBAL ATMOSPHERIC TURBULENCE DECISION SUPPORT SYSTEM FOR AVIATION

John K. Williams*, Robert Sharman, and Cathy Kessinger
National Center for Atmospheric Research, Boulder, Colorado

Wayne Feltz, Anthony Wimmers, and Kristopher Bedka
University of Wisconsin-Madison SSEC/CIMSS
1225 W. Dayton Street, Madison, WI 53706

1. INTRODUCTION

Turbulence is widely recognized as the leading cause of injuries to flight attendants and passengers on commercial air carriers. Oceanic and international routes are subject to clear-air turbulence, mountain wave turbulence (which for Atlantic flights is particularly significant over Greenland), and convectively-induced turbulence. Turbulence encounters may occur in remote regions where ground-based observations are sparse, making hazard characterization more difficult, and where international turbulence and convective SIGMETs provide only low temporal and spatial resolution depictions of potential hazards. Therefore, a new effort is underway to develop a global diagnosis and forecast system that will augment and enhance legacy products and provide authoritative global turbulence data for the World Area Forecast System and the Next Generation Air Transportation System's planned 4-D database of aviation weather data. This fully automated system, modeled on the FAA's Graphical Turbulence Guidance (GTG) and GTG Nowcast systems, will employ NCEP Global Forecast System model output and satellite data to produce quantitative turbulence nowcasts and forecasts. The convective nowcast methodology will make use of GFS data and operational data from GOES, Meteosat and MTSAT satellites, and will be tuned and verified using data from NASA's TRMM, Cloudsat and MODIS instruments. Satellite-based turbulence diagnosis algorithms will also be developed. AIREPs and AMDAR data will be used in conjunction with a machine learning methodology to develop an empirical model that maps the model fields, turbulence diagnoses and convective nowcasts and derived features to global deterministic and probabilistic nowcasts and forecasts of turbulence. This paper describes the first steps towards this system's development, including the use of artificial intelligence techniques for data analysis and algorithm development.

2. WORLD AREA FORECAST SYSTEM

The World Area Forecast System (WAFS) was established in 1982 by the International Civil Aviation Organization (ICAO) and the World Meteorological Organization (WMO) to provide weather guidance for international aircraft operations. In concert with the US Federal Aviation Administration (FAA), the National Weather Service (NWS) established the Washington World Area Forecast Center (WAFC) in 1997, which along with a second WAFC in London supplies worldwide aviation weather data. Unfortunately, current WAFS global weather products are issued infrequently and cover such large domains. For instance, the area enclosed by a significant meteorological information report (SIGMET) is typically so large that aircraft have little option but to traverse through it (see Figure 1). This is much like the situation over the conterminous United States (CONUS) until a few years ago when the Graphical Turbulence Guidance (GTG) product and the National Convective Weather Forecast (NCWF) became operational. GTG is an automated turbulence forecast system that produces 1, 2, 3, 6, 9 and 12 hour forecasts of turbulence intensity (Sharman et al. 2006); NCWF produces automated 0-2 hour forecasts of convection likelihood. Both products are used to support the creation of turbulence and convective AIRMETs and SIGMETs over the CONUS; they are also used to provide a graphical display directly to general aviation and commercial users (see http://weather.aero/). Clearly, global turbulence and convection decision support systems would benefit from an analogous global system, whose development is described in this paper. The "Global GTG" system will supply automated, comprehensive assessments of atmospheric turbulence for altitudes above 10,000 feet and at lead times of 0-36 hours.

3. IDENTIFYING SOURCES OF TURBULENCE

Aircraft turbulence encounters often occur in the vicinity of jet streams and upper-level fronts, near the tropopause, over mountainous regions, and in or near cloud, especially convective cloud (e.g., Hopkins 1977, Lester 1993, Chandler 1987, Sharman 2004, Wolff and Sharman 2007). Most of the clear-air sources are longer lived (e.g., Vinnichenko et al. 1980), but still may be rapidly evolving, and turbulence related to convection is known to be highly transient and spatially varying (e.g., Lane et al. 2003). Thus, Global GTG development is broken out into two primary components, one to diagnose longer-lived clear-air turbulence sources from GFS model output, and a second nowcast component that utilizes the most current observations to capture the more rapidly evolving features associated with mountain waves or convection.

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

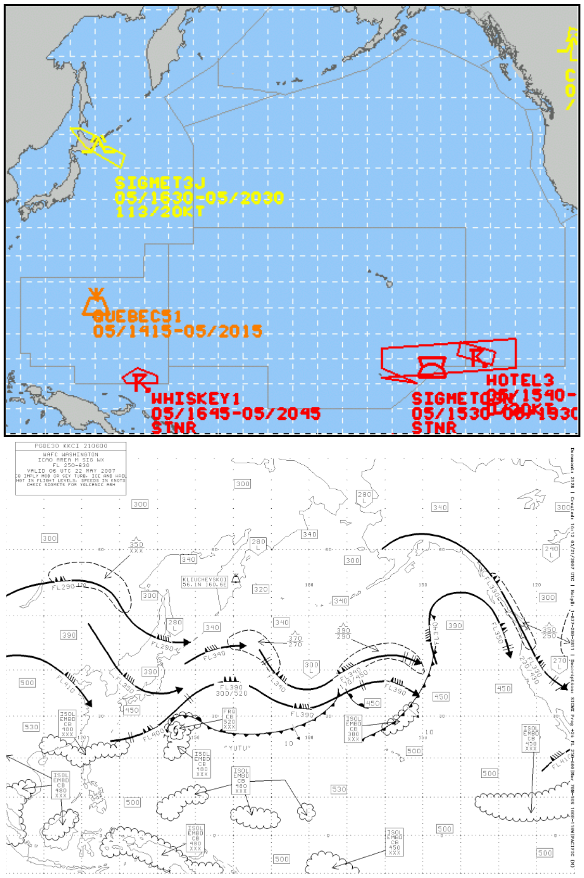

Figure 1: WAFS decision support products. (Top) Example graphical display of international SIGMETs, which are updated every 4 hours. (Bottom) Example of a SIGWX facsimile chart, updated every 6 hours.

A. Clear-air and mountain-wave turbulence

The Global GTG components for non-convective turbulence forecasts are modeled after the CONUS GTG (Sharman et al. 2006), which uses a suite of NWP model-derived diagnostics for clear-air turbulence (CAT) and mountain-wave turbulence (MWT), and only indirectly addresses turbulence associated with clouds and convection. In GTG, these diagnostics are combined in a weighted regression to get the best agreement with available observations. It was shown in Sharman et al. (2006) that a GTG combination based on static weights performs almost as well as a version where the weights are dynamically updated based on comparisons to recent PIREPS. Since the number of turbulence reports available at any one time globally will usually be insufficient for dynamic weighting, climatological weights are employed for Global GTG.

Adapting GTG to global use has required replacing the CONUS NWP model (currently RUC) with the GFS model to generate the constituent turbulence diagnostics globally; the combination of these diagnostics will then be tuned by training against *in situ* turbulence reports and AIREPs to optimize statistical performance globally. Figure 2 compares a GFS-based diagnostic to that generated from the RUC model. The overall patterns over the CONUS are quite similar, indicating that the coarser resolution GFS model is capable of capturing at least some the large-scale features conducive to turbulence formation. However, as the bottom set shows, some turbulence diagnostics appropriate for northern latitude break down at the equator. Thus, the selection of diagnostics and determination of appropriate combinations will have to be performed on a regional basis.

CAT may also be associated with tropopause folds, layers of stratospheric air that penetrate into the troposphere near a front and frequently exhibit dynamical instability (Shapiro 1980). Studies have shown that such folds may be identified via GOES water vapor gradients in conjunction with NWP model data, which establishes their vertical extent. The mechanism and example output from a prototype tropopause-fold detection algorithm (Wimmers and Moody 2004a, 2004b) is shown in Figure 3.

Turbulence may also be caused by mountain-waves created by the flow of air over rough terrain (Wurtele et al. 1996). Signatures of such waves are sometimes visible in satellite water vapor imagery (Uhlenbrock et al. 2005, 2007). Automated wave detection and characterization algorithms may provide an additional valuable contribution to the diagnosis of regions susceptible to MWT.

B. Convectively-induced turbulence

Convective or convectively-induced turbulence (CIT) may account for over 60% of turbulence encounters over the CONUS (Cornman and Carmichael 1993; see also Kaplan et al. 2005); the fraction may be

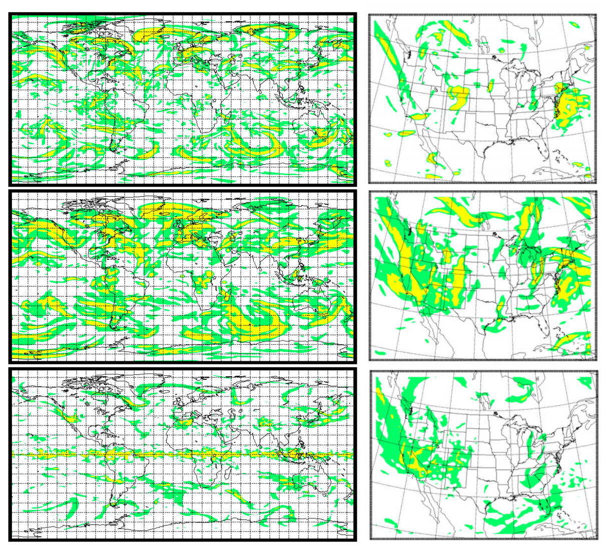

Figure 2: Comparison of turbulence diagnostics from the GFS 6-hour forecast valid at 18 UTC 4 November 2008 (left column) and RUC forecast (right column). The diagnostics are the Ellrod Index for 35,000 ft (top), EDR Index for 35,000 ft (middle) and Richardson number from thermal wind at 20,000 feet (bottom).

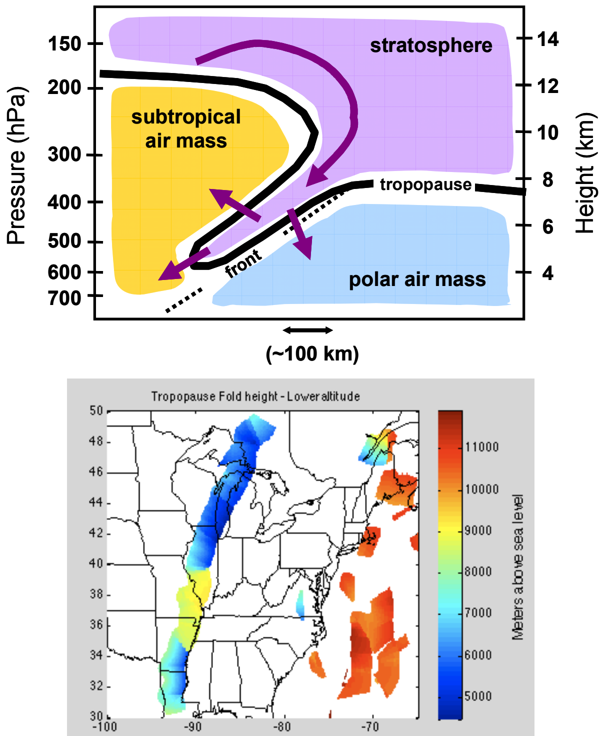

Figure 3: Illustration of the tropopause-fold mechanism (top) and sample output from a prototype satellite and model-based tropopause-fold detection algorithm with color-coded altitudes (bottom).

somewhat less in oceanic regions due to typically high cruise altitudes and fewer routing constraints, but remain a significant concern. Convection can develop and evolve rapidly, and it may not be well predicted by NWP models. For this reason, it is necessary to augment the GFS forecasts with assessments of CIT based on observation-based convection analyses and nowcasts. An algorithm for diagnosing CIT over the CONUS based on a combination of environmental conditions from the RUC model and thunderstorm features from satellite IR values and trends, radar reflectivity, radar turbulence detection, and lightning data was described in Williams et al. (2007). Developing an analogous global CIT diagnosis system requires utilizing GFS model data in place of RUC and modifications to accommodate diverse geographical domains and the more limited storm characterization data available in oceanic and remote regions.

A combination of the GFS model plus satellite data has been developed to characterize the presence of convection and environmental conditions likely to support convective development (Kessinger et al. 2009). Storms are identified and characterized based on a cloud-top height derived from satellite longwave infrared data and GFS model temperature profiles. The Global Convective Diagnosis (GCD; Mosher 2002) utilizes the difference between the satellite-measured longwave and water-vapor channels to identify deep convective clouds that have reached the tropopause. The cloud top and GCD are combined to obtain a scalar metric of thunderstorm intensity (see Figure 4). Thunderstorm nowcasting may be achieved via storm extrapolation, or a technique that involves data fusion of observation and model data (Cai et al. 2009). Convective nowcasting products originally developed for use with the US GOES and Japanese MTSAT-1R satellites are being modified for use with Meteosat data to obtain coverage of Africa, Europe and Asia. NASA TRMM, CloudSat, and CALIPSO data, along with CONUS radar reflectivity and lightning data, are being used to evaluate the convective products' performance. Convection locations and intensity will be used in combination with GFS model fields to diagnose regions of potential CIT.

In addition to the location and intensity of convection, additional storm features may be useful for diagnosing CIT. For instance, using satellite longwave infrared data, locations of "overshooting tops" where a developing storm penetrates the stratosphere may be identified by looking for pixels that are cold and substantially colder than their surroundings.

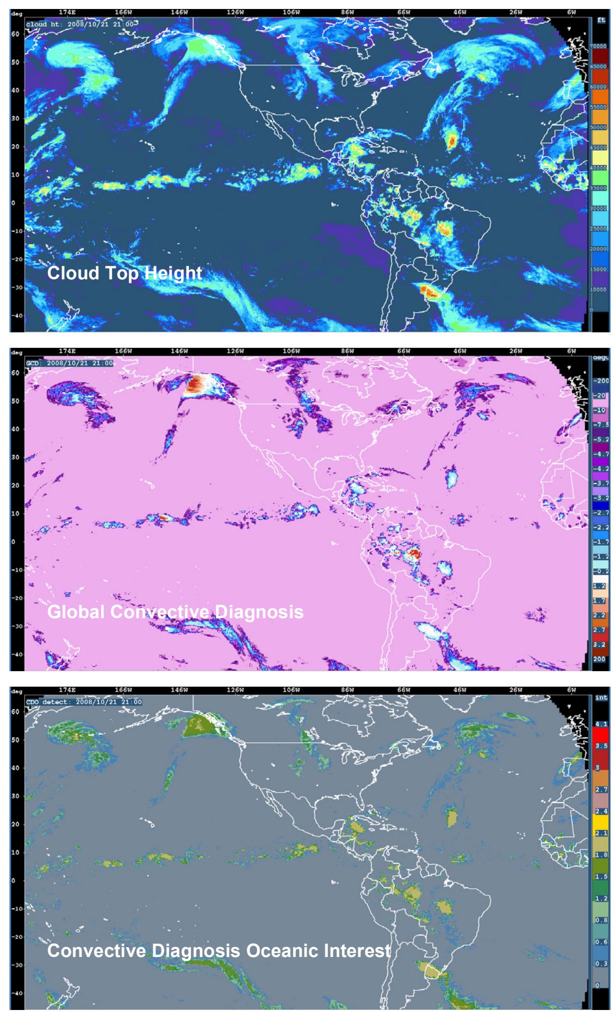

Figure 4: Satellite-derived convection information include cloud-top height (top), Global Convective Diagnosis (middle), and a convective intensity metric called the Convective Diagnosis Oceanic (bottom).

4. AI ANALYSIS AND DATA FUSION APPROACH

A machine learning technique called random forests (RFs; Breiman, 2001) has shown promise in analyzing potential predictor fields and developing an empirical model. RFs have previously proven useful in developing empirical models for diagnosing regions of CIT (Williams et al. 2007) and for developing thunderstorm nowcasts (Williams et al. 2008a, Cai et al. 2009). Essentially, an RF is a group of decision trees which collectively form an "ensemble of experts" that "vote" on the correct classification of an input. In the course of training an RF, information is obtained on the relative importance of various candidate predictor variables; this new information may in turn be useful in understanding the underlying physical process or identifying distinct phenomenological "regimes" (Williams et al. 2008b, Abernethy 2008). The RF approach has proven to work well even when several hundred predictors require analysis.

In order to use the RF methodology, it is first necessary to establish a training dataset in which various data instances are associated with the desired classification label, or "truth". Over the CONUS, Hawaii, and Central and northern South America, automated *in situ* reports from United and Delta Airlines aircraft provide routine measurements of EDR, a quantitative turbulence metric (Cornman et al. 1994, 2004) that may be used for this purpose. However, these data are not available globally. Sources of oceanic and international turbulence information include qualitative air reports (AIREPs) and measurements of U_{de} (maximum derived vertical gust, based on true air speed fluctuations) from the automated Aircraft Meteorological Data Relay (AMDAR) system. The geographical distribution of these measurements is shown in Figure 5.

These available turbulence "truth" data cover a variety of conditions and locations, which is important since both turbulence phenomenology and the quality of predictor fields are expeted to vary geographically (recall the failure of the bottom field in Figure 2 near the equator). Unfortunately, the geographic coverages of available sources of turbulence measurements do not overlap significantly, making it very difficult to compare and inter-calibrate them. Moreover, there are some geographic areas, such as South America and Africa, that have very little or no coverage. This creates a significant challenge to a machine learning technique in developing a consistent global turbulence product.

The approach that has been developed to address this problem is twofold. First, each of the available truth datasets will be used independently in evaluating predictor importance and developing data fusion empirical models. The *models* will then be calibrated to one another by comparing their output on their common domains of applicability; then they will be combined to form the global forecast. Second, in developing the empirical models, a first step will be to identify "regions" in which similar combination logic is appropriate. These regions may be functions of the surface type (land/ocean, terrain, surface temperature, climate, or circulation), latitude, and season. Appropriate regions may be determined by training a random forest over local areas and combining those areas having similar predictor importance rankings, as was done for CONUS turbulence forecasting in Abernethy (2008). Once the attributes that determine the regions are identified, the same logic may be extended to similar areas in which no truth data are available. In short, the attributes that affect the turbulence phenomenology will be learned so that models can be appropriately generalized.

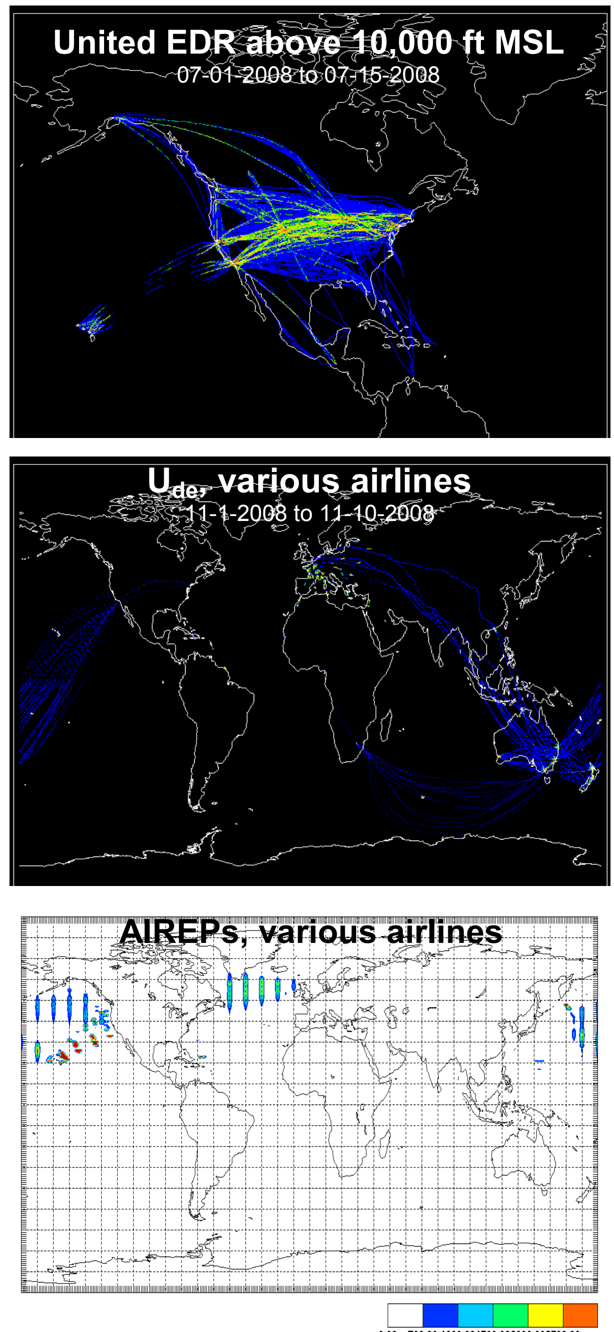

Figure 5: Geographical distribution of turbulence "truth" data sources: *In situ* EDR reports (top), U_{de} measurements (middle), and AIREPs (bottom).

The final result of this algorithm development approach will be a combination of empirical models that provide integrated 3-D forecasts and nowcasts of atmospheric turbulence. These will include 0-3 hour nowcasts that address CAT, MWT and CIT and 3-36 hour model-based forecasts. Separate models will be created for each forecast lead time.

Performance evaluation and tuning of the turbulence forecasts and nowcasts will be based on both qualitative measures and quantitative techniques. Qualitative assessments by demonstration users of the turbulence and convection nowcasts will include United Airlines pilot responses to a questionnaire that asks them to describe the flight conditions and the accuracy and utility of each experimental cockpit uplink message received relative to standard weather information sources. This information will be collected via a website.

Quantitative assessments of the turbulence forecasts and nowcasts will be performed using two primary assessment metrics. The first performance measure is based on receiver operating characteristic (ROC) curves (e.g., Marzban 2004, Sharman et al. 2006). Sharman et al. (2006) compared ROC curves for individual turbulence diagnostics and the GTG combination against AIRMETs and showed that GTG performed better than any individual turbulence diagnostic, and the improvement in GTG's accuracy over the operational AIRMETs was substantial. Similar accuracy improvements of Global GTG over international SIGMETs are expected.

A second performance measure used in evaluations of probabilistic forecasts is the reliability diagram (e.g., Wilks 1995, Hamill 1997), in which the predicted probabilities of turbulence over some threshold (light, moderate or severe) are compared with the measured relative frequency of turbulence encounters given that prediction. A well-calibrated forecast or nowcast will produce a reliability curve close to the 1:1 line. This approach also accounts for forecast volumes, which should be as small as possible. Both the reliability diagram and ROC curves will be computed from all available data in the global domain for as long a period as is practical, and the turbulence product will be tuned to optimize its performance.

5. SUMMARY AND FUTURE WORK

This paper has described the motivation and technical elements involved in the development of Global GTG, a software system that will use GFS model data and data from several satellites along with intermediate products that address various sources of turbulence to generate comprehensive gridded, probabilistic turbulence forecasts and nowcasts and convection nowcasts. Global GTG will contribute to improving WAFS products, and will supply global turbulence information for the NextGen 4-D aviation weather data cube. A significant challenge in the development of Global GTG will be utilizing the disparate and geographically-distinct truth datasets available for training empirical data fusion models. This challenge will be addressed through a combination of regionalization and comparison and calibration of the models generated from the various truth data sources.

Global GTG will be demonstrated via a web-based graphical display and cockpit uplinks of customized convection and turbulence information to selected commercial flights. This will allow qualitative evaluation via pilot feedback in addition to quantitative verification based on comparisons with *in situ* turbulence data and NASA satellite convection data.

6. ACKNOWLEDGEMENTS

This research is based upon work supported by NASA under Grant No. NNX08AL89G. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration.

Some of this research was performed 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.

Several software engineers at NCAR/RAL have contributed significantly to this project. They include Gary Blackburn, Jason Craig, Greg Meymaris, and Nancy Rehak.

7. REFERENCES

Abernethy, J., 2008: *A Domain analysis approach to clear-air turbulence forecasting using high-density in-situ measurements*. Ph.D. Dissertation, Department of Computer Science, University of Colorado.

Breiman, L., "Random forests," Machine Learning, **45**, 5-32 (2001).

Cai, H., C. Kessinger, D. Ahijevich, J. Williams, N., Rehak, D. Megenhardt, R. Bankert, J. Hawkins, M. Donovan, and E. Williams, 2009: Nowcasting oceanic convection using random forest classification, *AMS 16th Conference on Satellite Meteorology and Oceanography, Fifth Annual Symposium on Future Operational Environmental Satellite Systems-NPOESS and GOES-R, and Fourth Symposium on Policy and Socio-Economic Research*, paper J8.4.

Chandler, C. L., 1987: Turbulence forecasting. In *Atmospheric Turbulence Relative to Aviation, Missile, and Space Programs*, NASA Conference Publication 2468, 137-154.

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. *J. 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. *AMS Eleventh Conf. on Aviation, Range, and Aerospace Meteorology*, P4.3.

Hamill, T. M., 1997: Reliability diagrams for multicategory probabilistic forecasts. *Wea. Forecasting*, 12, 736-741.

Hopkins, R. H., 1977: Forecasting techniques of clear-air turbulence including that associated with mountain waves. WMO Technical Note No. 155, 31 pp.

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.

Kessinger, C., H. Cai, N. Rehak, D. Megenhardt, M. Steiner, R. Bankert, J. Hawkins, M. Donovan, and E. Williams, 2009: The oceanic convection diagnosis and nowcasting system. *AMS 16th Conference on Satellite Meteorology and Oceanography, Fifth Annual Symposium on Future Operational Environmental Satellite Systems-NPOESS and GOES-R, and Fourth Symposium on Policy and Socio-Economic Research*, paper J8.3.

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.

Lester, P. F., 1993: *Turbulence: A new perspective for pilots*. Jeppesen Sanderson, Inc., 212 pp.

Marzban, C., 2004: The ROC curve and the area under it as performance measures. *Wea. Forecasting*, 19, 1106-1114.

Mosher, F., 2002: Detection of deep convection around the globe. *Preprints, 10th Conf. Aviation, Range, and Aerospace Meteor., Amer. Meteor. Soc.,* Portland, OR, 289-292.

Shapiro, M. A., 1980: Turbulent mixing within tropopause folds as a mechanism for the exchange of chemical constituents between the stratosphere and troposphere. *J. Atmos. Sci ,* 37*,* 994-1004.

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

Sharman, R.D, 2004: Clear-air Turbulence (CAT). *Encyclopedia of Nonlinear Science*, Alwyn Scott, Ed., 1104 pp.

Uhlenbrock, N., S. Ackerman, W. Feltz, T. Whittaker, B. Sharman, K. M. Bedka, and L. Gumley, 2005: The use of satellite water vapor imagery and model data to diagnose and forecast turbulent mountain waves. *21st International Conf. on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography and Hydrology,* American Meteorological Society, P2.40.

Uhlenbrock, N. L., K. M. Bedka, W. F. Feltz, and S. A. Ackerman, 2007: Mountain wave signatures in MODIS 6.7 um imagery and their relation to pilot reports of turbulence. *Wea. Forecasting*, 22, 662-670.

Vinnichenko, N. K., N. Z. Pinus, S. M. Shmeter, and G. N. Shur, 1980: *Turbulence in the Free Atmosphere.* Plenum, 310 pp.

Williams, J. K., D. Ahijevych, S. Dettling and M. Steiner, 2008a: Combining observations and model data for short-term storm forecasting. In W. Feltz and J. Murray, Eds*., Remote Sensing Applications for Aviation Weather Hazard Detection and Decision Support. Proceedings of SPIE*, 7088, paper 708805.

Williams, J. K., D. A. Ahijevych, C. J. Kessinger, T. R. Saxen, M. Steiner and S. Dettling, 2008b: A machine-learning approach to finding weather regimes and skillful predictor combinations for short-term storm forecasting. *AMS 6th Conference on Artificial Intelligence Applications to Environmental Science and 13th Conference on Aviation, Range and Aerospace Meteorology*, paper J1.4.

Williams, J. K., 2007: "Introduction to Fuzzy Logic." A chapter to appear in *Artificial Intelligence Applications in Environmental Science*, S. E. Haupt, C. Marzban and A. Pasini, Eds. In press.

Williams, J. K., J. Craig, A. Cotter, and J. K. Wolff, 2007: A hybrid machine learning and fuzzy logic approach to CIT diagnostic development. *AMS Fifth Conf. on AI Applications to Env. Science*, 1.2.

Wimmers, A. J. and J. L. Moody, 2004a: Tropopause folding at satellite-observed spatial gradients: 1. Verification of an empirical relationship, *J. Geophysical Research – Atmospheres*, 109, art. no. D19306.

Wimmers, A. J. and J. L. Moody, 2004b: Tropopause folding at satellite-observed spatial gradients: 2. Development of an empirical model, *J. Geophysical Research – Atmospheres*, 109, art. no. D19307.

Wolff, J. K., and R. D. Sharman, 2007: Climatology of upper-level turbulence over the continental United States. Submitted to the *J. of Appl. Meteor. Climo.*

Wurtele, M. G., R. D. Sharman, and A. Datta, 1996: Atmospheric lee waves. *Annu. Rev. Fluid Mech.,* 28,429-76.