An Artificial Intelligence Approach to Operational Aviation Turbulence Forecasting

Jennifer Abernethy
National Center for Atmospheric Research
University of Colorado
Boulder, CO
aberneth@ucar.edu

Robert Sharman
National Center for Atmospheric Research
Boulder, CO
sharman@ucar.edu

Elizabeth Bradley
University of Colorado
Boulder, CO
lizb@cs.colorado.edu

Abstract— Turbulence is a major aviation hazard for both commercial and private aircraft. Currently, the clear-air turbulence forecasting tool Graphical Turbulence Guidance (GTG) is used by airline meteorologists and dispatchers for flight planning, and in part to determine operational Airman's Meteorological Information (AIRMET) turbulence advisories; however, GTG has much higher resolution and intensity discrimination than do AIRMETs, providing more pinpointed locations of moderate or greater turbulence. Because numerical weather prediction (NWP) models cannot explicitly predict aircraft-scale turbulence, we use artificial intelligence (AI) algorithms to capture the relationships between large-scale atmospheric conditions and turbulence. This paper provides an overview of GTG and details beginning work for development of the next release of GTG using in-situ turbulence observation data. We apply two AI techniques, support vector machines and logistic regression, to clear-air turbulence prediction. We show improved forecast accuracy over the current product performance, and begin specializing forecasts by geographic region and altitude. We show the algorithms' feasibility as part of a real-time operational turbulence forecasting system.

I. Introduction

Pilots' ability to avoid turbulence during flight affects the safety of the millions of people who fly commercial airlines and other aircraft every year. Turbulence is a rare event in terms of percentage of the atmosphere that is turbulent at any given time [10], however, of all weather-related commercial aircraft incidents, 65% can be attributed to turbulence encounters, and major carriers estimate that they receive hundreds of injury claims and pay out "tens of millions" per year [26].Turbulence can occur in and around thunderstorms, over mountains, near the ground, in clouds, and even in clear air. At upper levels, clear-air turbulence, or CAT, is particularly hard to avoid because it is invisible to traditional remote sensing techniques. The dynamical scales in which CAT appears, however, are far finer than those of any current weather prediction model. And observations of the state of the system – reports of 'light' or 'moderate/severe' radioed in by pilots who encounter turbulence – are sparse and subjective.

In 1998, the Federal Aviation Administration (FAA) Aviation Weather Research Program (AWRP) funded the National Center for Atmospheric Research Research Applications Lab (NCAR/RAL) to develop a graphical decision support tool, now called Graphical Turbulence Guidance (GTG), which provides clear-air turbulence (CAT) forecasts over the continental U.S. (CONUS). GTG became operational in 2003. Meteorologists and dispatchers at the major airlines have access to GTG forecasts through the National Weather Service (NWS) Aviation Weather Center's (AWC) Aviation Digital Data Service (ADDS) website to use in planning and altering flight routes. AWC forecasters consider GTG forecasts when producing Airman's Meteorological Information (AIRMET) turbulence advisories, also available on ADDS. Future development plans include merging CAT forecasting with other forecasting products and weather information in the Joint Product Development Office (JPDO) Next Generation Air Transportation System (NextGen), a comprehensive four-dimensional weather information source for aviation decision support (information available online at http://jpdo.gov)

The turbulence forecasting difficulty is due to two main factors: (1) turbulent eddies at the scales that affect aircraft (~100m) are a microscale phenomenon and operational numerical weather prediction (NWP) models cannot resolve that scale (NWP models that are run operationally by the National Weather Service, for instance, which produce hourly and daily weather predictions, only capture what's happening every 10 to 20 km) and (2) lack of objective observational turbulence data. The prior factor has been addressed during the past 50 years, by assuming that most of the energy associated with turbulent eddies at aircraft scales cascades down from larger scales of atmospheric motion [9,20,28].The turbulence forecast problem then becomes one of linking large-scale features resolvable by NWP models to the formation of aircraft-scale eddies. Numerous 'rules of thumb' empirical linkages, termed turbulence *diagnostics*, were developed by the National Weather Service, airline meteorologists and academic researchers. The forecast skills of these diagnostics depend on the forecaster (for manual forecasts) and diminish with lead time. The diagnostics' skills reflect in part researchers' imperfect understanding of the atmospheric processes involved.

The second problem is being addressed by a new, better source of turbulence observations, termed *in-situ data*. In-situ data are sensor data from aircraft: measures of atmospheric eddy dissipation rate (Cornman et al., 2004). While the study of CAT is necessarily limited to that directly experienced by aircraft since it cannot be seen, in-situ data is so much more

ISBN: 978-0-615-20720-9

plentiful than PIREP observations that researchers now have enough data to explore additional AI techniques for forecasting.

This paper details the atmospheric and observational data used in turbulence forecasting, the current operational algorithm, and how new AI approaches, used both over the entire domain and regionally, in combination with new, more plentiful in-situ turbulence observations to improve operational CAT forecasting.

II. Background

A. Turbulence Diagnostics

Through the years when forecasts were done manually, forecasters developed “rules of thumb” about what atmospheric conditions typically indicated turbulence. These rules of thumb were an attempt to link the available large-scale meteorological information to the micro-scale CAT that was the subject of the forecast [13]. Forecasters later quantified these rules, creating CAT diagnostics. A CAT diagnostic is a simple turbulence model (equation) calculated from NWP model data. For instance, a major cause of CAT is the Kelvin-Helmholtz instability: when gravity waves become steep and unstable, they may break into a chaotic motion [9]. This typically happens in areas of strong vertical shear (the difference in velocity between horizontal layers) and low local Richardson number (Ri, the ratio of static stability and wind shear), so many CAT diagnostics involve shear and Ri. There are many different diagnostics, each linking a large-scale condition to small-scale turbulence. Their predictive powers vary, depending upon the large-scale condition that each represents and how directly it is linked to turbulence. A full explanation of the forty CAT diagnostic equations can be found in [26].

Forecasters use these diagnostics by mapping their values to different turbulence severity levels. As an example, low Ri indicates high turbulence. Early on, forecasters determined some unofficial thresholds to quantify the severity of turbulence that corresponded to a given diagnostic value--- “Ri < 0.25 = moderate or greater turbulence," for example [9]. In this way forecasters were able to transform their qualitative knowledge to a quantitative form (diagnostics) which could be used in automated systems. GTG developers used several years’ worth of PIREPs to develop threshold values for each diagnostic that map to different levels of PIREP turbulence severity. This allows the diagnostics to work neatly with the qualitative PIREP observations. Since there are many problems with PIREP accuracy, and optimal thresholds may change depending on the day or season, we hope to avoid this step in the next forecasting system. We expect that AI techniques will do this thresholding step within their algorithm and thus can either find the best thresholds themselves, and/or respond to the dynamic relationship between large-scale and small-scale atmospheric processes by adjusting these thresholds during training.

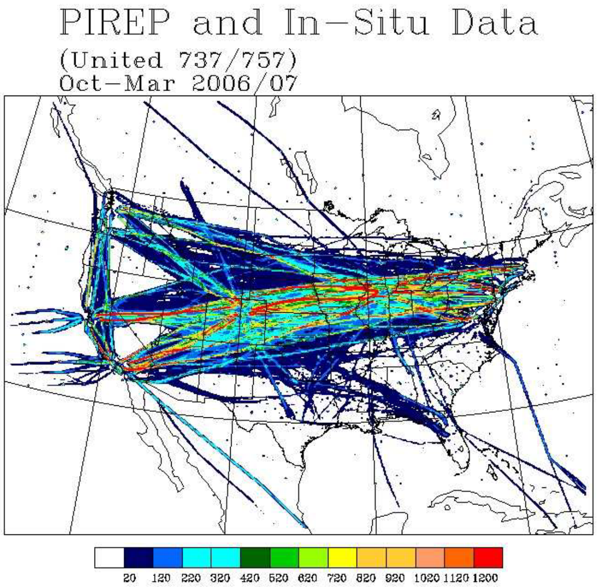

Figure 1. The PIREP and in-situ data used in this study, showing the geographic distribution of in-situ data. PIREP data are all but invisible under dense in-situ data along United flight paths; some can be seen as points in the southeast and surrounding the U.S. Color indicates frequency of observations.

B. In-Situ Data

In-situ turbulence measurements are sensor data that are recorded by special software on commercial aircraft every minute during flight. Detailed coverage of in-situ data methods can be found in [6] and [7]. Specifically, in-situ measurements are an estimate of atmospheric turbulence intensity called the eddy dissipation rate (EDR) around an aircraft. Eddies are irregular currents of air, and the rate at which eddies break down is recognized as a good measure of atmospheric turbulence intensity [23]. Compared to PIREPs, in-situ data are more objective, more accurate, more plentiful, and more representative of the actual distribution of turbulence in the atmosphere [8, 26]. At any one time, over 99% of the atmosphere is expected to be free of turbulence [10].

Currently, in-situ measurements of EDR are being gathered from 197 United Airlines aircraft. Several other airlines will deploy the same system in the coming year. The in-situ data used in this study, from October-March 2006/7, is shown in Figure 1. Each in-situ data report is a location triple (latitude, longitude, altitude) and a median and peak (95th percentile) EDR reading from measurements taken over the corresponding minute. Each of the two EDR fields is binned and the two binned values are transmitted to the ground. The binning turns otherwise continuous quantitative observation data into a set of eight discrete values that are cognate to the eight PIREP intensity levels. Currently, we consider bin 4 to correspond to a ‘moderate’ PIREP, although study to better qualitatively understand in-situ data is ongoing [2].

C. Performance Metrics

It is not trivial to assess the accuracy of a forecast because we do not know the ‘truth’; we must use available observation data, however flawed or irregular. We follow the verification practices of [27], which include the Receiver Operating Characteristic (ROC) curve and area under the ROC curve (AUC) [22], and True Skill Score (TSS). A ROC curve measures how well an algorithm discriminates between Moderate or Greater (MOG) and less than moderate (LTM) turbulence. To construct the curve, we vary the threshold that separates these two classes over a range of 0 to 1 and measure the discrimination accuracy at each threshold. Two numbers are used to capture this: PODY, "probability of detecting a yes" (forecast made a correct positive (MOG) prediction), and PODN, which corresponds to a correct negative (LTM) prediction. Higher the PODY/PODN combinations over the range of thresholds implies greater classification accuracy, so the area under the curve (AUC) is a useful single-number metric for forecast accuracy. The TSS considers PODY and PODN at one threshold: TSS = PODY + PODN – 1. As mentioned in subsection B, our threshold for this study is in-situ bin 4 and 'moderate' PIREPs: bin 4 and above constitute the MOG category, and below bin 4 constitutes the LTM category.

D. Graphical Turbulence Guidance System

The GTG forecasting product produces a graphical display of turbulence severity for each flight level, FL100 to FL450, over the CONUS, for zero, six, nine and 12 hour forecasts, updated every hour. Displays of the operational product (GTG1) are available in real-time on the ADDS website, http://adds.aviationweather.noaa.gov. The newer version, GTG2, is available on the experimental ADDS website, http://weather.aero. An example is shown in Figure 2, with the AIRMET forecast for comparison of forecast specificity.

Every hour, the algorithm calculates ten diagnostics from the National Center for Environmental Prediction's (NCEP) Rapid Update Cycle (RUC) model at 20km resolution for that hour [3]. These diagnostics are paired with incoming PIREPs from a time window around the RUC time, usually 1.5 hours. A fuzzy logic technique scores each diagnostic based on its agreement with the observation data, deriving a set of weights such that the weighted sum of the diagnostic values is between zero and one. The scoring function, per diagnostic over the CONUS, incorporates TSS and percentage of the CONUS forecast as MOG turbulence (f_{MOG}):

$$\phi_n = \left(\frac{TSS + 1.1}{1 + Cf_{MOG}} \right) \qquad (1)$$

Currently, C=1. From the *n* scores (one for each diagnostic over the entire forecast area), weights are formed:

$$W_n = \phi_n{}^2 \qquad (2)$$

subject to $\sum_{m=1}^{n} W_m = 1$. The diagnostics are combined into a weighed sum to form the GTG combination, an estimated turbulence intensity for every grid point:

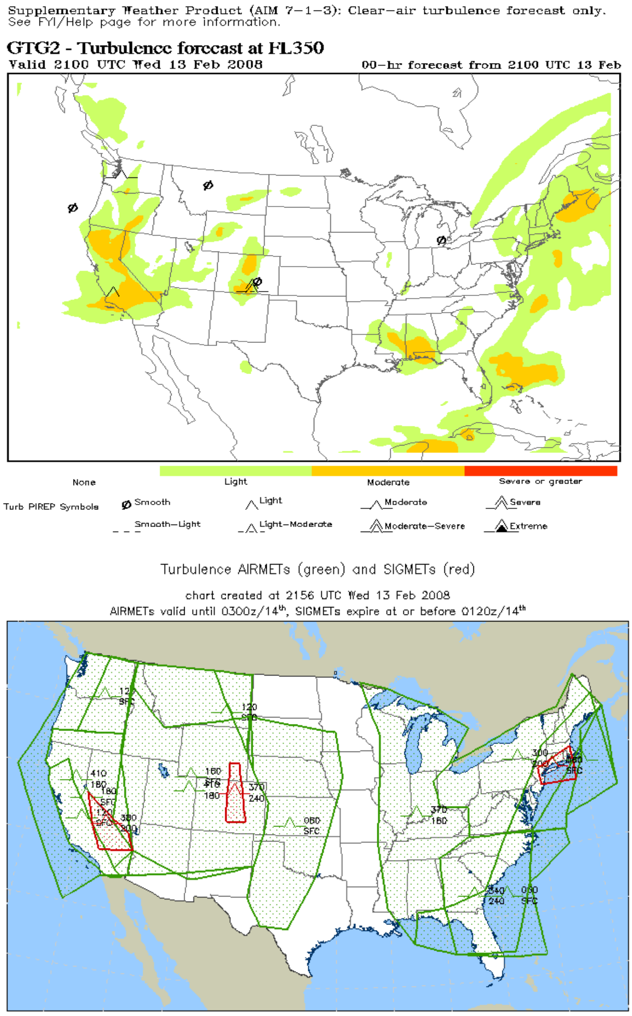
Supplementary Weather Product (AIM 7–1–3): Clear–air turbulence forecast only.
See FYI/Help page for more information.
GTG2 - Turbulence forecast at FL350
Valid 2100 UTC Wed 13 Feb 2008
00–hr forecast from 2100 UTC 13 Feb
None
Light
Moderate
Severe or greater
Turb PIREP Symbols
Smooth
Light
Moderate
Severe
Smooth–Light
Light–Moderate
Moderate–Severe
Extreme
Turbulence AIRMETs (green) and SIGMETs (red)
chart created at 2156 UTC Wed 13 Feb 2008
AIRMETs valid until 0300z/14th, SIGMETs expire at or before 0120z/14th
120
SFC
410
180
300
370
240
080
SFC

Figure 2. An example of a GTG2 forecast display (above) which designates light, moderate, severe and extreme CAT per hour for each flight level, and AIRMET (below), in green, which show much larger polygon areas of moderate turbulence for a corresponding 6-hour period over all flight levels. Both displays are from http://weather.aero.

$$GTG(i, j, k) = \sum_{m=1}^{n} W_m D_{m,i,j,k} \qquad (3)$$

The set of weights is applied to the RUC diagnostics calculated from the 6, 9 and 12 hour RUC forecast. This dynamic weighting allows GTG to respond to changing conditions every hour. While the fuzzy logic handles sparse PIREP observation data well, the new, plentiful in-situ data allows for more choices in prediction algorithms.

E. Artificial Intelligence Techniques

Generally, a classifier is an algorithm that predicts a data classification given (presumably) relevant data features. The Support Vector Machine (SVM) is a popular machine learning technique for classification. The SVM produces a model that predicts the class label by setting parameter values of an optimization problem based on its input data [15]. Here, class labels are MOG turbulence, and LTM turbulence (see subsection C).

In order to learn the relationships (parameter values) between these data features and the class label, we first train a classifier by giving it many known feature/class pairs. Each pair in the training set is known as a data instance. A data instance k consists of a set of features $x_{i,k}\ \ i = 1...n$ and a target class label *y*.

During training, each feature vector X_k is mapped into a higher dimensional space. The SVM finds a linearly separating hyperplane with the maximal margin between class means in this higher dimensional space. To classify an example, the SVM calculates the distance of that example to each class mean through a series of dot products, and classifies it in whatever class has the closest mean [5]. This series of dot products is at the heart of the model and is a measure of vector similarity called a kernel function:

$$K(x_i, x_j) = \phi(x^T{}_i)\phi(x_j) \tag{4}$$

For implementation of the SVM, we use the LibSVM library [4]. LibSVM provides four basic kernels and an optional program, "grid.py", which selects the model (i.e., does a parameter search). Previous studies [1,2] show good performance for the radial basis function kernel:

$$K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2), \gamma > 0 \tag{5}$$

The radial basis function kernel only has two parameters: γ and *C*, a penalty parameter for the SVM error term. A model can output probabilities for class membership, also.

A logistic regression equation is solved iteratively over the training set, determining a prediction equation with a parameter, or weight, for each feature. The response variable, between zero and one, is the log odds that the class label is one. We can interpret the log odds as a probability, and decide that *P(y) >=0.5* should be classified as MOG. Background on logistic regression can be found in [14].

III. Methodology

Our initial application of AI techniques to operational turbulence prediction consisted of testing Support Vector Machine (SVM) and logistic regression algorithm performance over our entire prediction domain, the CONUS. For each algorithm, for both zero-hour and six-hour forecasts, we used a subset selection search to pick a subset of CAT diagnostics which together had the highest forecast accuracy. We then tested the performance of each model in a simulated operational real-time system using either a static model for each hour's forecast or dynamic training of the model using the previously-chosen subset. We've taken the first steps to make specific forecasts, with specific sets of diagnostics, for different regions of the CONUS. The following subsections summarize our data and methods.

A. Data

Data used in the current development stage consist of weather model and observation data – both PIREPs and in-situ data – from October through March 2005/6 and 2006/7, shown in Figure 1. The weather model is the RUC NWP model at 13km resolution, run operationally and disseminated every hour by the National Center for Environmental Prediction (NCEP) [3]. RUC model data was used to calculate forty CAT diagnostics for each RUC model grid point and observation data was matched by time and location to the forty diagnostics for a grid point. Since we are primarily interested in predicting CAT, an upper-level phenomenon, only matches above 20000ft were used.

The distribution of the data used during the training process is a very important factor in the ability of a classifier to discriminate between the two classes [17]. The winter 2006/7 data set, as an example, contained nearly nine million observation/diagnostics matches. Over 98% were LTM turbulence. SVMs, for instance, aim for the lowest overall error rate, and could simply classify everything as LTM and have a less than 2% error. This is well-supported in the literature [16, 29, 30]. To work well, the training data set must have a large number of examples from each class; we rebalanced the training data such that 40% of the data were of MOG, and 60% were LTM. We did this by keeping all the MOG observations and choosing LTM observations randomly to be 60% of the set. This proportion of MOG/LTM resulted in the best SVM classification rate in an earlier study of SVMs with CAT diagnostics and in-situ data [1]. We found 20% MOG and 80% LTM to be a good distribution for logistic regression training data.

Analysis of the data reveals that PIREPs dominate the MOG category (>92%) – partly due to inadequate special coverage of in-situ data at this time – and in-situ data dominates the LTM category (>98%). Thus, PODY is effectively a measure of the algorithm's ability to predict PIREPs and PODN is a measure of in-situ prediction capability. Since in-situ data is more objective, we know using only in-situ data to train the algorithm improves performance [2]. However, our forecasting product will be verified using PIREPs, at least in part, and they cover more geographical area than do in-situ data, thus we cannot abandon them yet.

B. Subset Selection Search

Turbulence forecasting, in its current state, is essentially the task of classifying atmospheric indicators of turbulence: the forecast reflects the number of diagnostics which indicate turbulence in an area. While it might seem obvious to simply use the individually best-performing diagnostics for forecasting, as was done with GTG, that approach allows one to possibly miss a different set of diagnostics that might perform better, as a group, than the set of the individually top-ranked diagnostics [12, 19, 20]. Our search for the best subset of diagnostics is essentially the task of feature subset selection [12]. We are faced with the choice between 40 diagnostics, knowing that some may not improve our current forecasting accuracy. In addition, it is infeasible to calculate and use all 40 in a real-time operational system. The wrapper method in feature subset selection executes a state space search for a good feature subset, estimating prediction accuracy using an induction algorithm – here, we used SVMs or logistic regression [19], with TSS as a final scoring metric. We used a simple hillclimbing search. Each state is a subset of diagnostics, and the search operator is "add a diagnostic". The search chooses the best addition to the current subset based on the classification performance of the induction algorithm using the current subset plus an additional diagnostic. This approach to the search is called forward selection. Thus, we start with an empty subset and added diagnostics stepwise; our stopping condition was no further classification performance improvement (measured by no change in TSS). Searches were performed for SVM and logistic models for both zero and six-hour forecasts using training, testing and holdout data sets from 18Z over winter 2006/7.

C. Simulated Real-Time System

We have created a simulated real-time forecasting system capable of using either SVMs or logistic regression to create a turbulence forecast every hour for the CONUS (like GTG). The system is capable of training a model for every forecast hour or using a pre-trained model so that we may test performance differences between dynamic and static weighting, respectively. For both, we use the sets of diagnostics found in the searches explained in subsection B. Results are from trials over the fifteen day period of 2/1/2007 to 2/15/2007. Thus far, we have concentrated on zero-hour forecasts in this step.

We had to take several steps to make the SVM algorithm feasible for a real-time system. Since LibSVM uses ASCII files, 13km-resolution gridded RUC data caused each forecast to take over an hour. To streamline processing, we built a NetCDF file format interface onto the library. We also replaced the exponential function with an approximation. Both changes cut the forecast time down to a more operationally-appropriate five minutes.

D. Regionalization

Thus far, we have conducted regionalization studies using SVMs only, on winter 2005-2006 data. We employed subset searches for each of these regions: west of 100W meridian, east of 100W, above and below 30000ft (20000 to 30000ft), and by both geography and altitude (e.g., east of 100W and below 30000ft: low east). We plan to further refine and divide regions in the near future, but for this study, we have simply isolated the mountainous terrain in the west region from the non-mountainous terrain in the east.

TABLE I. AREA UNDER THE CURVE, TRUE SKILL SCORE, AND SUBSET SIZE RESULTS FOR FEATURE SELECTION SEARCHES AND 0-HR 15-DAY REAL-TIME SIMULATION RUNS. GTG SKILLS FOR THE SAME DATA ARE IN ITALICS. HIGHER TSS AND AUC INDICATE GREATER SKILL.

	AUC	TSS	Subset size
GTG 0hr fcsts	*0.795*	*0.390*	*10*
Log search 0hr	0.801	0.478	13
SVM search 0h	0.7825	0.471	8
GTG 6hr fcsts	*0.78*	*0.366*	*10*
Log search 6hr	0.79	0.467	6
SVM search 6h	0.78	0.4643	12
GTG 0-hr 15days	*0.799*	*0.350*	*10*
Log static 0-hr	0.823	0.466	13
SVM static 0-hr	0.796	0.459	8
Log dyn. 0-hr	0.786	0.45	13
SVM dyn. 0-hr	0.775	0.464	8

TABLE II. INITIAL REGIONALIZATION RESULTS USING SVM FOR WINTER 2005/6. THE GTG TSS FOR THIS PERIOD WAS 0.453; EVEN ARBITRARY REGIONALIZATION SHOWS IMPROVEMENT WITH SPECIALIZED SETS OF DIAGNOSTICS.

Region	TSS	Set of Diagnostics
West	0.465	6
East	0.562	4
>=30000ft	0.447	5
<30000ft	0.607	5
High west	0.441	4
High east	0.516	4
Low west	0.614	5
Low east	0.519	2

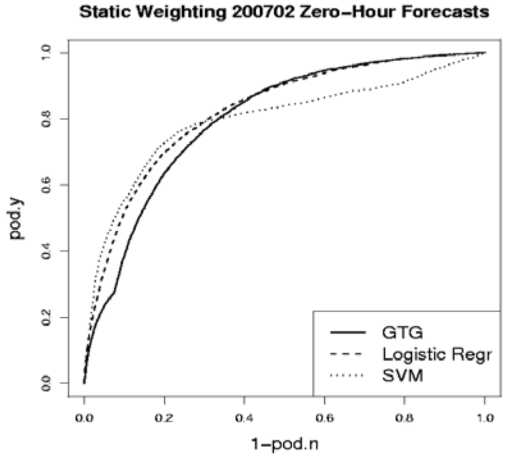

Figure 3. Receiver operating characteristic (ROC) curves comparing performance for 15-day real-time simulation of 0-hr forecasts using static weighting. The solid line is the current GTG performance for the same 15-day period. Lines closer to top left corner indicate better forecasting performance. See Table 1 for areas under the curves.

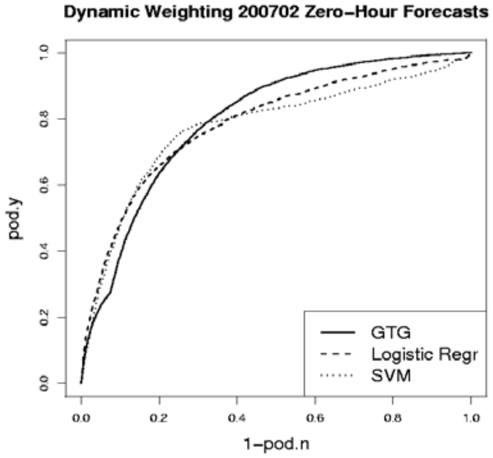

Figure 4. Like Figure 3, for dynamically-trained models.

ISBN: 978-0-615-20720-9

IV. Results

Results of our forward selection subset searches and real-time simulations are shown in Table 1. While we do not list the exact diagnostics chosen by each search for the sake of simplicity, we did find that there was significant — though not complete — overlap in the diagnostics chosen by each search, indicating high predictive capability for a core subset of four or five diagnostics. Logistic regression shows a small improvement in AUC over the overall performance of the current GTG algorithm for both 0 and 6-hr forecasts (about a 0.01 difference), however, the true-skill scores (TSSs) for both algorithms are significantly improved over GTG (0.09 – 0.1 improvement). This is most likely due to the fact that our search used TSS as the heuristic to choose the sets of diagnostics.

Figures 3 and 4 show the ROC curves for our static- (model trained in the search step is applied to data from each hour) and dynamic-weighting (new model is trained every hour) 15-day real-time simulations. It should be noted that GTG has been tuned using years of PIREPs, thus its PODY scores are highest (since PIREPs dominate the PODY category). Logistic regression using pre-determined (static) weights improves significantly upon the current GTG product, increasing the AUC from 0.799 to 0.823 and the TSS from 0.350 to 0.466. While the static-weighting SVM and both dynamically-weighted models had similar improvements in TSS over GTG, we saw no improvement in AUC (mainly due to reduced prediction skill for the MOG category). TSS is discrimination skill at the MOG threshold, 0.375; AUC measures classification skill at many thresholds. Thus, we have improved forecasting performance at the operational MOG threshold, although the ROC curves show us that there is still need for improvement in the algorithms overall.

Our initial regionalization results are shown in Table 2. For winter 2005/2006, the baseline TSS for the GTG algorithm was 0.453; thus, most regions showed improvement in forecasting accuracy when using chosen subsets of diagnostics. In addition, the fact that different diagnostics were chosen in the different regions indicate that diagnostics can perform differently in different areas of the country, reflecting the geographic differences in the large-scale atmospheric processes they represent and good potential for the regionalization approach.

V. Conclusions

Forecasting clear-air turbulence is critical to aviation safety. AI techniques can be very useful in meeting the challenges inherent in this process because they smoothly handle sparse, noisy data sets, significant levels of uncertainty, and gaps in the understanding of the underlying physical mechanisms, all of which are characteristics of the turbulence prediction domain.

This paper has detailed the first steps in applying the artificial intelligence techniques of support vector machines and logistic regression to clear-air turbulence forecasting. While the current GTG product uses fuzzy logic, the algorithmic choices were limited by the sparse PIREP observation data; now, the more objective and plentiful in-situ data vastly widens the choices for prediction algorithms. We've shown not only improvement in forecasting performance both globally and regionally, but also the feasibility of implementing these AI algorithms in a real-time operational product setting. Future work includes continued study of these algorithms for regionally-specific forecasting and probabilistic forecasting.

Acknowledgment

This research is in response to requirements and funding by the FAA. The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA.

References

[1] Abernethy, J., 2005: Domain Analysis Approach to Clear-Air Turbulence Forecasting Using In-situ Data. Dissertation Proposal, Department of Computer Science, University of Colorado.

[2] Abernethy, J., Bradley, E.; and Sharman, R. 2006: Qualitative reasoning about small-scalle turbulence in an operational setting. In Proceedings of the Qualitative Reasoning Workshop. Hanover, NH.

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

[4] Chang, C. and C. Lin. LIBSVM – a library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[5] Chen.P., C. Lin and B. Scholkopf, 2003: A tutorial on v-support vector machines. http://kernel-matchines.org.

[6] 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.

[7] Cornman, L., G. Meymaris, and M. Limber, 2004: An update on the FAA Aviation Weather Research Program's in-situ turbulence measurement and reporting system. Preprints, *Eleventh Conf. on Aviation, Range, and Aerospace Meteorology*, Hyannis, MA, Amer. Meteor. Soc., P4.3.

[8] Dutton, J., 1980: Probability forecasts of clear-air turbulence based on numerical output. Meteor. Mag., **109**, 293-310.

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

[10] Frehlich, R., and R. Sharman, 2004a: Estimates of turbulence from numerical weather prediction model output with applications to turbulence diagnosis and data assimilation. *Mon. Wea. Rev.*, 132, 2308-2324.

[11] Frehlich, R., and R. Sharman, 2004b: Estimates of upper level turbulence based on second order structure functions derived from numerical weather prediction model output. Preprints, *Eleventh Conf. on Aviation, Range and Aerospace Meteorology*, Hyannis, MA, Amer. Meteor. Soc., P4.13.

[12] Guyon, I. and A. Elisseef, 2003: An introduction to variable and feature selection. *J. Machine Learning Research*, 3, 1157-1182.

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

[14] Hosmer, D. and S. Lemeshow.1989. Applied Logistic Regression. John Wiley and Sons, Inc.

[15] Hsu, C., C. Chang and C. Lin, 2003: A practical guide to support vector classification. Published online with Libsvm documentation at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[16] Japkowicz, N., 2000: Learning from imbalanced data sets: a comparison of various strategies. *AAAI Workshop on Learning from Imbalanced Data Sets*, Menlo Park, CA.

[17] Kay, M., J. Henderson, S. Krieger, J. Mahoney, L. Holland and B. Brown, 2006: Quality assessment report: Graphical turbulence guidance (GTG) version 2.3.

[18] Kohavi, R., and D. Sommerfield, 1995: Feature subset selection using the wrapper method: overfitting and dynamic search space topology. *First International Conference on Knowledge Discovery in Data Mining (KDD-95).*

[19] Kohavi, R. and G. John, 1997: Wrappers for Feature Subset Selection. *J. Artificial Intelligence*, **97**, *no1-2, 273-324.*

[20] Koshyk, J. N., and K. Hamilton, 2001: The horizontal energy spectrum and spectral budget simulated by a high-resolution troposphere-stratosphere-mesosphere GCM. *J. Atmos. Sci ,* **58**, *329-348.*

[21] Kronebach, G. W., 1964: An automated procedure for forecasting clear-air turbulence. *J. Appl. Met.,* **3,** 119-125.

[22] Marzban, C. 2004: The ROC Curve and the Area Under It as a Peformance Measure. *Weather and Forecasting*, Vol. 19, No. 6, 1106-1114.

[23] Panofsky, H and J. Dutton, 1983: Atmospheric turbulence: models and methods for engineering applications. John Wiley & Sons.

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

[25] Sharman, R., G. Wiener and B. Brown, 2000: Description and verification of the NCAR integrated turbulence forecasting algorithm. *Proceedings of the 38th Aerospace Sciences Meeting and Exhibit, Reno, NV.*

[26] Sharman, R., C. Tebaldi, G. Wiener and J. Wolff, 2006: An Integrated Approach to Mid- and Upper-Level Turbulence Forecasting. *Weather and Forecasting*.

[27] Takacs, A., L. Holland, R. Hueftle, B. Brown and A. Holmes, 2005: Using in-situ eddy dissipation rate (edr) observations for turbulence forecast verification.

[28] Tung, K. K., and W. W. Orlando, 2003: The k^{-3} and $k^{-5/3}$ energy spectrum of atmospheric turbulence: Quasigeostrophic two-level model simulation. *J. Atmos. Sci*., **60**, 824-835.

[29] Weiss, G. and F. Provost, 2001: The effects of class distribution on classifier learning: an empirical study. *Technical Report ML-TR-44, Department of Computer Science, Rutgers University.*

[30] Wu, G and E. Chang, 2005: KBA: Kernel boundary alignment considering imbalanced data distribution. *IEEE Transactions on Knowledge and Data Engineering.*

ISBN: 978-0-615-20720-9