Probabilistic Turbulence Prediction Using Random Forests

Zhengzheng Li[1,2,4] zhengzheng.li@ou.edu

Timothy A. Supinie[3,4] Timothy.A.Supinie-1@ou.edu

[1]Atmospheric Radar Research Center (ARRC)
[2]School of Electrical and Computer Engineering
[3]School of meteorology
[4]University of Oklahoma, Norman, Oklahoma 73019-6151, USA

Abstract

The turbulence prediction problem for the AI contest contains measurement from aircraft, satellite, ground-based radar and numerical weather model. According to the peak EDR values, cases can be classified into two categories: moderate-or-greater turbulence or no turbulence. Random Forest has been widely used on redundant and weakly correlated dataset and it is adopted in this study. Unlike normal Random Forest having each tree in the forest to vote, a new technique of combining results from the forest is proposed to generate probabilistic forecast, which is the AI contest requirement. Quality of each variable is tested by randomly varying its values and measuring the performance difference before and after the variation. Small portion of variables are then selected based on their qualities and used in Random Forest.

1. Introduction

Studies have shown that turbulence in and around thunderstorms, which is also known as convectively-induced turbulence (CIT), causes over 60% of turbulence-related aircraft accidents. Accurate turbulence prediction algorithm that deals with turbulence especially CIT can significantly improve airplane safety.

There exist many studies on turbulence detection in literature. Relation between radar reflectivity and turbulence is discussed in [1] where optimum radar frequency for clear-air turbulence (CAT) detection is derived. It is also demonstrated in this paper that reflectivity in excess of some minimum threshold value is a sign of some degree of turbulence. Also, results of simultaneous studies of turbulence in the lower 15 km of the atmosphere by multi-wavelength radar, jet aircraft and special rawinsondes show that radar is more sensitive to turbulence between altitudes of 3 and 6 km ([2]). How turbulent air motion contributes to mean and variance of radar spectrum and what kind of Doppler radar spectrum indicates turbulence are discussed in [3]. Besides radar, satellites are very common tools used in turbulence detection. From satellite image analysis, scientists show that CIT are often found in association with (1) Rapidly vertical convective development; (2) Rapidly expanding anvil clouds indicative of strong outflow or divergence; (3) Banded cirrus outflow structures (i.e. transverse bands); (4) Convective gravity waves ([4]).

Unlike many current turbulence detection/prediction algorithms, which are primarily based on physical knowledge of how turbulence is formed, the proposed algorithm in this study focuses on how to predict turbulence from measured data in a statistical way regardless what those data are. However, by identifying the quality of each variable in the data, this statistical approach helps better understand turbulence.

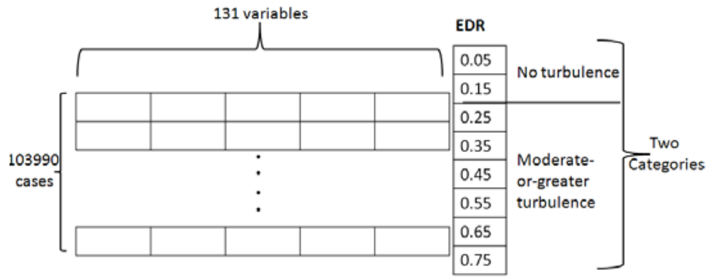

Figure 1 - Problem Definition

2. Problem Definition and Algorithm

2.1 Task Definition

Dataset used in this paper is from AI contest website. They are mainly collected during summer months in which convectively-induced turbulence (CIT) is particularly prevalent, though mountain-wave turbulence (MWT) and clear-air turbulence (CAT) are also present. In this dataset, collocated observation and model-derived variables have been extracted for each aircraft EDR measurement. More specifically, observations from satellite, radar and simulated weather field from NWP model surrounding the plane's EDR measurement location have been used to calculate potential predictor variables that may have skill individually or in combination to indicate turbulence.

The dataset contains of 103990 cases at different time and locations. Each case has 131 variables (features). However, there are many cases where the satellite or radar readings are missing or null; those field values are labeled 'NA' in the data set. According to EDR measurement from an airplane, there are eight classes with binned EDR values 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65 and 0.75. Values of 0.25 or greater are considered as moderate-or-greater (‘Turbulence detected’), otherwise ‘No turbulence’, as shown in Figure 1.

2.2 Algorithm Definition

After random forest was introduced in [5], it has been implemented in many applications where there are redundant and weakly-correlated variables. It has been proven to be particularly well-suited in turbulence detection application by [6-9] where Random Forest is adopted and used to detect turbulence or help identifying important variables for further classification.

Random Forest make use of an ensemble of decision trees as described in [5]. Decision trees are attractive classifiers because they are easy to train and the execution speed of them is high. However, trees derived with traditional methods often cannot be grown to arbitrary complexity for possible loss of generalization accuracy on unseen data. Random Forest addresses this problem by growing multiple trees in randomly selected subspaces of the training dataset. In other words, each tree (randomized decision tree) in the forest is grown as follows:

1. Randomly select $2/3$ cases from training dataset.
2. When determining which input feature to split on at each node, only a random small subset of the input features is considered eligible.
3. Each tree is grown to the largest possible extent. There is no pruning.

A Random Forest is simply a set of some amount of randomized decision trees. Patterns that are truly present in the training data would be learn by most of the trees, while those that are simply relics of the sampling would be learn by fewer trees. Given a new input, this algorithm puts the input vector down through each of the trees in the forest and then chooses the class having the most votes (over all the trees in the forest).

In this study, missing data in training data set are filled with zero, though there is no significant difference in performance when filled with the mean value of that particular variable. 200 - 400 trees are grown in different experiments. Only 12 random features are selected at each node. Performance gets worse when this number is too small. However, no significant difference when this number is large enough. The output of the forest, which is the predicted probability of turbulence, is calculated by the weighted average of the votes (0 or 1) from all trees in the forest as given in equation (1).

$$P(turbulence) = \frac{\sum W_i Vote_i}{\sum W_i} \qquad (1)$$

W_i is the weight of ith tree. It relates to the Brier Skill Score of this tree and it is obtained from equation (2).

$$W_i = \frac{1}{1 - BSS_i} \qquad (2)$$

3. Experimental Evaluation

3.1 Methodology

For a probabilistic forecast to be reliable, the frequency of an observed event, should agree with the forecasted probability value (long term). However, a forecast with perfect resolution will always correctly forecast either 0% or 100% (short term). As required by the AI contest, the output probability of turbulence should reward both reliability and resolution, thus will be assessed using the Brier Skill Score (BSS). The Brier Skill Score combines features of resolution, reliability and observational uncertainty. The reliability component of the Brier Skill Score is the standard deviation of the difference between the forecast probability and the average frequency of the observed value corresponding to that forecast. This component should be minimized. The resolution component is the variance of the difference between the climatological frequency of an event occurring and the individual forecasts. This value should be maximized. In this turbulence prediction application, assume the ratio of turbulence in the dataset is p, therefore BSS can be calculated from equation (3).

$$BSS = \frac{p(1-p) - MSE}{p(1-p)} \qquad (3)$$

where MSE is the mean square error of the forests. As shown in equation (3), $BSS = 0$ means no skill and $BSS = 1$ means perfect skill.

To further assist evaluation of the forests, ROC plots are also used.

Since in Random Forest, each tree is trained by a different random subset of the original training data, there is no need for a separate test set to unbiased estimate of the test set error. After a tree is constructed, out-of-bag data (data that does not use in training) is used to test the current tree. Results (votes) are kept until all trees in the forest are finished. Performance of the forest can be obtained by evaluating those results.

To measure variable importance, values of each variable have been randomly permuted. Performance difference in experiments before and after the permutations is used to determine the quality of this particular variable. Two types of difference are measured. One is decrease in BSS and the other one is importance, which is decrease in correct votes.

3.2 Experiments

Mainly three kinds of experiments have been performed:

1. Use all variables and try to maximum BSS.
2. Use variables from different sources and try to figure out which source provides the best indications of turbulence.
3. Use small portion of variables that have better quality and try to figure out what variables are needed in turbulence prediction.

3.3 Results and Discussion

In the first experiment, as shown in Figure 2, a Random Forest with 400 trees is constructed on all variables. For each individual tree in the forest, BSS value is around -0.15 which is worse than no skill. However, by combining results from several trees, BSS value quickly increases to around 0.37 and then slowly increases as the number of trees further increase. Figure 2(b) shows the learning curve of this Random Forest. If keep increasing the number of trees, BSS value won't increase much and will decrease after certain amount of trees. This is caused over fitting. The forest is further evaluated using the ROC plot, as shown in figure 2(c). The performance of this agent is good.

In the second experiment, as shown in table 1, qualities of measure data from different sources are investigated. Variables extracted from satellite measurement are the worse feature while variables from simulated weather field are the best feature. If there's no other source besides radar in an airplane, this algorithm still performs very well.

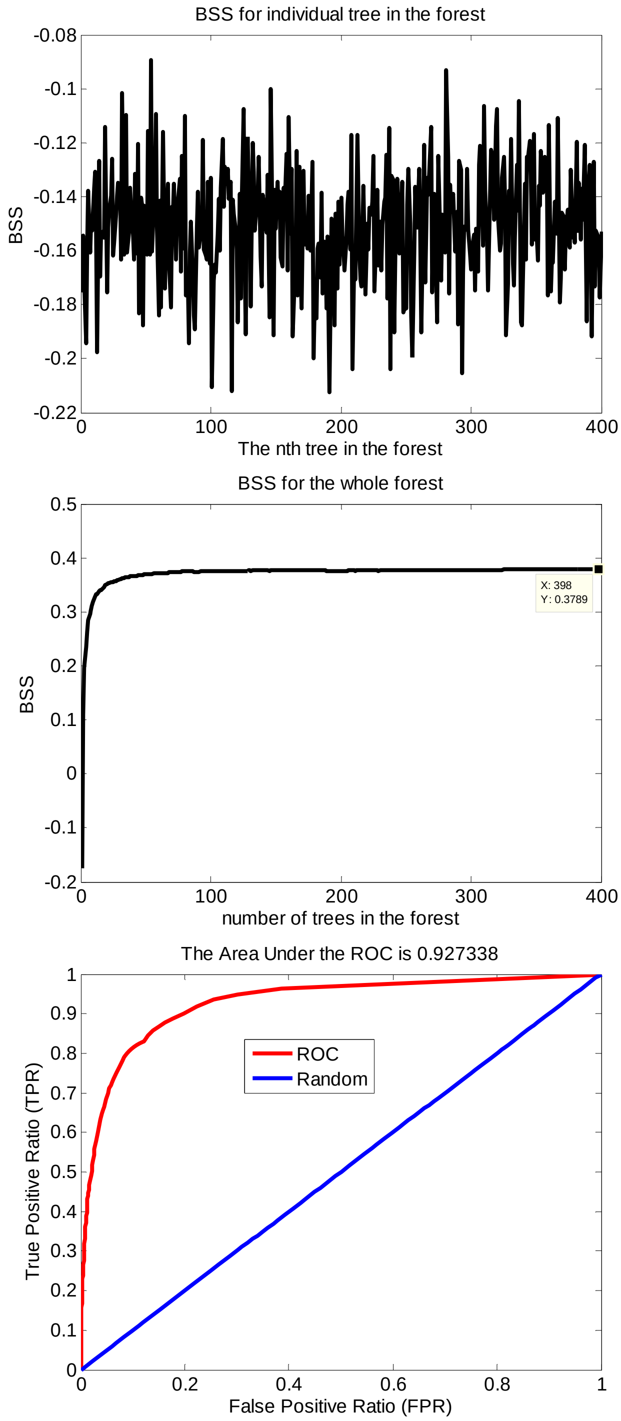

Figure 2(a),(b),(c) - Results for 400 trees

	Accuracy	ROC	BSS
Airplane	0.938	0.83	0.184
Satellite	0.937	0.64	0.076
Radar	0.943	0.88	0.289
NWP	0.946	0.91	0.332
AP + Radar	0.947	0.91	0.352
All	0.951	0.927	0.379

Table 1 - Performances of measurements from different sources

In the third experiment, as shown in Figure 3 and Table 2, quality of each variable is tested. Some variables are significantly better than other variables. If only use those good quality variables, will the performance improve? The answer is no. This may indicate that Random Forest is able to tolerant certain amount of noise. Though using just some good variables doesn't improve performance of this algorithm, it doesn't ruin the performance either. If only small portion of variables are allowed in such turbulence prediction algorithm, variables with better quality can be selected.

4. Related Work

Random Forest and fuzzy-logic are the key approaches in a series of papers ([6-9]) from National Center for Atmospheric Research (NCAR) where several kinds of turbulence detection algorithms are describes. Random Forest is used to identify important variables from measurements and then a fuzzy logic algorithm is built on those variables. Accuracy of their system is around 80% for eight categories.

Results from the Graphical Turbulence Guidance System (GTG) and support vector machines (SVM) are compared in [10, 11], where SVM is shown better than GTG in terms of True Skill Score (TSS).

It is very difficult to compare the proposed algorithm in this study with other approaches. Because training dataset, input variables and target categories are different. However, in terms of ROC and BSS, the performance of this proposed algorithm is good.

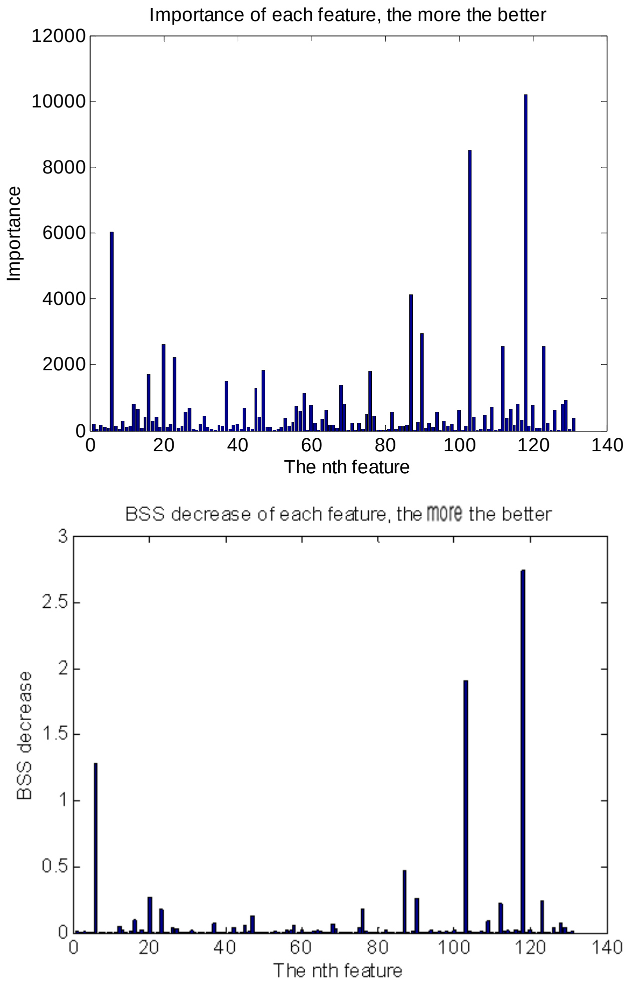

Figure 3(a),(b) - Quality of features

	10%	20%	30%	40%	50%
Importance	0.3	0.342	0.358	0.368	0.373
BSS decrease	0.31	0.349	0.357	0.363	0.367

Table 2 - Performance (BSS) when use different percentages of importance/BSS decrease features

5. Future Work

The current shortcoming of this turbulence detection algorithm is that there are no velocity information used. As shown in many studies, mean and variance of the Doppler velocity spectrum are two very good indicators of turbulence. In future work, velocity information should be included.

Some variables may have their own distributions and can be statistically modeled. Further exploration of the variables is needed.

6. Conclusion

A random forest approach with good performance in turbulence detection/prediction is discussed. Quality of variables from different source is tested. Using small portion of good quality variables cannot improve the performance but slightly impair it. Turbulence detection using measurement from aircraft and radar is also investigated and proven to be feasible.

References

[1] D. Atlas, K. R. Hardy, and K. Naito, "Optimizing the Radar Detection of Clear Air Turbulence," *Journal of Applied Meteorology*, vol. 5, no. 4, pp. 450–460, 1966.

[2] K. M. Glover, R. J. Boucher, H. Ottersten *et al.*, "Simultaneous Radar, Aircraft and Meteorological Investigations of Clear Air Turbulence," *Journal of Applied Meteorology*, vol. 8, no. 4, pp. 634–640, 1969.

[3] A. R. Bohne, "Radar Detection of Turbulence in Precipitation Environments," *Journal of the Atmospheric Sciences*, vol. 39, no. 8, pp. 1819–1837, 1982.

[4] K. M. Bedka, W. Feltz, J. Mecikalski *et al.*, "Satellite signatures associated with significant convectively-induced turbulence events," *13th Conference on Aviation, Range and Aerospace Meteorology,New Orleans, LA*, 2008.

[5] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

[6] A. Cotter, J. K. Williams, R. K. Goodrich *et al.*, "A Random Forest Turbulence Prediction Algorithm," *Fifth Conference on Artificial Intelligence Applications to Environmental Science,San Antonio, TX*, 2007.

[7] J. K. Williams, L. Cornman, D. Gilbert *et al.*, "REMOTE DETECTION OF TURBULENCE USING GROUND-BASED DOPPLER RADARS," *11th Conference on Aviation, Range, and Aerospace,Hyannis, MA*, 2004.

[8] J. K. Williams, J. Craig, A. Cotter *et al.*, "A HYBRID MACHINE LEARNING AND FUZZY LOGIC APPROACH TO CIT DIAGNOSTIC DEVELOPMENT," *Fifth Conference on Artificial*

Intelligence Applications to Environmental Science,*San Antonio, TX*, 2007.

[9] J. K. Williams, R. D. Sharman, and C. J. Kessinger, “Developing a global atmospheric turbulence decision support system for aviation,” *Seventh Conference on Artificial Intelligence and its Applications to the Environmental Sciences*,*Phoenix, AZ*, 2009.

[10] J. Abernethy, R. Sharman, and G. Wiener, “TOWARDS A REGIONAL CLASSIFIER APPROACH TO AVIATION TURBULENCE PREDICTION,” *13th Conference on Aviation, Range and Aerospace Meteorology*,*New Orleans, LA*, 2008.

[11] J. A. Abernethy, and R. Sharman, “Support vector machines for regional clear-air turbulence prediction,” *Fifth Conference on Artificial Intelligence Applications to Environmental Science*,*San Antonio, TX*, 2007.