8th Conference on Artificial Intelligence Applications to Environmental Science

Session 3.1

The Third Annual Artificial Intelligence Forecasting Competition

M. J. Pocernich , NCAR, Boulder, CO; and Stephen Sullivan and J. Abernethy

A forecast that includes an expression of uncertainty or a probability can be much more useful than the forecast of a single value. For this reason, the 2009 AMS Artificial Intelligence Competition was based on forecasting a probability. The challenge for the 2009 contest was to predict the probability of turbulence exceeding a specific threshold.

1. Convectively Induced Turbulence (CIT)

Contestants were asked to predict the chance that atmospheric turbulence would exceed a threshold that could affect aviation. Data used in the contest was collected during summer months – from June through September. This is the period when CIT--turbulence in and around thunderstorms--is particularly prevalent. Some mountain-wave turbulence (MWT) and clear-air turbulence (CAT) were also present. Studies have suggested that CIT is responsible for over 60% of turbulence-related aircraft accidents; thus, accurate real-time turbulence diagnoses that include CIT could 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 atmospheric turbulence, and CIT in particular, are a topic of current research and are still only partially understood. However, the likelihood of CIT is thought to be related to the proximity (vertical and horizontal), intensity, depth and extent of convection as well as the state of the atmosphere around the storm. It seems plausible that an empirical model that uses numerical weather prediction model data to get an indication of larger-scale environmental conditions along with satellite, radar reflectivity, and lightning observations that indicate the extent and severity of the storms and associated clouds could have good skill in predicting turbulence. MWT could similarly be modeled based on location (e.g., the presence of rough topography) as well as environmental conditions reflected in numerical weather prediction model data, and CAT may be predicted based on environmental conditions. Observations of turbulence generally entail pilot reports or automated reports. Automated reports of eddy dissipation rate (EDR, a measure of atmospheric turbulence) produced every minute by a collection of commercial aircraft were used as the observed values for turbulence for this contest.

2. Evaluating Probabilistic Forecasts

There are two important attributes for probabilistic forecasts - reliability and resolution. For a probabilistic forecast to be reliable, the frequency of an observed event, should agree with the forecasted probability value. For example, when a forecast of 20% is made, one should observe this event 20% of the time. When this is true, a forecast is considered reliable. However, a reliable forecast is not necessarily a useful forecast. By only forecasting the long-term chance of an event occurring, one would have a reliable forecast, but one can readily see this has limited utility. For this reason, one also needs to consider the resolution of a forecast. A forecast with perfect resolution will always correctly forecast either 0% or 100%. A completely random forecast or a completely consistent forecast such as the climatological average probability has no resolution.

To reward both reliability and resolution, the forecasts in this competition was 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. This is done when forecasts are either 0% or 100% in correct proportions to the climatological frequency.

The Brier Skill Score is not without its weaknesses. The value of the BSS, like all skill scores, is dependent on the sample climatology. Different climatologies will result in different scores. In this competition, everyone used the same sample dataset, so comparing scores is appropriate. Also, with these two components, a single Brier Skill Score can be the result of different combinations of resolution and reliability components. In the reality, different uses will have specific requirements for resolution and reliability. Further information about the Brier Score can be found at Jolliffe and Stephenson (2003), Wilks (2006) or W WWRP/WGNE (2010).

3. Contest Data

The training dataset contained 103,990 data rows, and 136 columns including the binary response variable of peak edr values exceeding a threshold. The test dataset contained 50,127 data rows in a format similar to the training dataset, but without the response variable.

3.1 Response Variable

The object of this contest is to predict the probability that the measured turbulence is moderate-or-greater (MoG). The response variable to be predicted is 0 (false) if the EDR measurement reflects null or light turbulence, and 1 (true) if it is above the threshold for MoG turbulence.

3.2 Predictor Variables

Potential predictors were co-located observation and model-derived variables, extracted for each aircraft EDR measurement. The NWP model, satellite and radar fields surrounding the plane's EDR measurement location were used to calculate potential predictor variables that indicate a plane's distance from various intensity levels of storms and clouds, as well as environmental characteristics at the measurement point. Since this was real data, many times that the satellite or radar readings are missing. Since MoG turbulence is quite rare, the proportion of null to positive instances in both training and testing datasets has been manipulated for the purposes of this contest by removing 2/3 of the null report instances.

A summary of the variables used in the contest is provided below.

- Airplane information at time the EDR measurement.
- Aircraft id, time and location
- Lightning information:
- Satellite radiance channels from the NOAA GOES imager:
- NEXRAD radar-derived storm intensity and proximity information:
- NWP model-derived fields:
- Rapid Update Cycle (RUC) numerical weather prediction model analysis. The values were linearly interpolated from the model grid to the location of the EDR measurement.

4. Results

Figure 1 shows the results of the Brier Skill Scores. Confidence intervals were estimated by re-sampling the data, with replacement, by date. For each re-sample, the same number of dates are used, with some dates are represented multiple times while others are omitted. This method was used based on the assumption that conditions within a single date are more strongly correlated than between dates. With the BSS, larger values indicate a better forecast. So for these forecast, contestant Li Zhengzheng (#4) was clearly the contest winner. A proper pairwise comparison between entrants was not conducted.

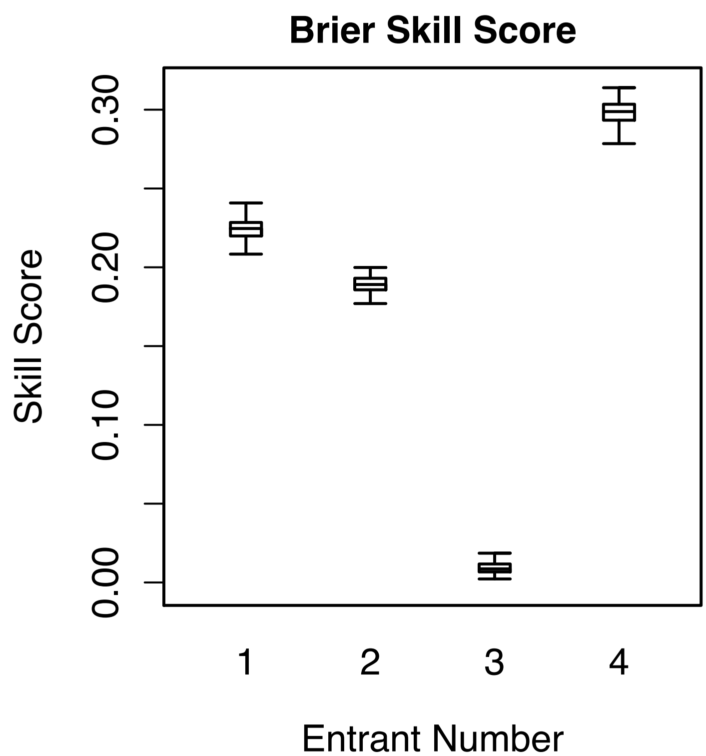

Figure1. Distribution of Brier Skill Scores from bootstrapped test data.

Table 1 shows the BSS values for each contestant and the title of their presentation delieverd at the AMS meeting in Atlanta, 2010. Li Zhengzheng from the Univsersity of Oklahoma was the winner. Congratulations to all who participated.

Table 1. Brier Skill Scores for predictions on test dataset.

Id #	Brier Skill Score Median (5^{th},95^{th} CI)	Presentation Title	Authors
1	0.22, (0.21, 0.23)	Turbulence Probability using Principal Component Analysis and Support Vector Machine Approaches	**Kimberly L. Elmore**, CIMMS/Univ. of Oklahoma and NOAA/NSSL, Norman, OK; and M. B. Richman
2	0.18, (0.18. 0.20)	Statistical Turbulence Prediction	**Walter C. Kolczynski Jr.**, Penn State University, University Park, PA; and S. E. Haupt
3	0.1, (0.00, 0.02)	Predicting Turbulence Using a Neural Network	**Valliappa Lakshmanan**, CIMMS/Univ. of Oklahoma, NOAA/NSSL, Norman, OK
4	0.29 (0.28, 0.31)	Probabilistic Turbulence Prediction using Random Forests	**Zhengzheng Li**, The University of Oklahoma, Norman, OK; and T. A. Supinie

References

Jolliffe, I.T., and D.B. Stephenson, 2003: Forecast Verification. A Practitioner's Guide in Atmospheric Science. Wiley and Sons Ltd, 240 pp.

Wilks, D.S., 2006: Statistical Methods in the Atmospheric Sciences. 2nd Edition. Elsevier, 627 pp

WWRP/WGNE Joint Working Group on Forecast Verification Research (2010), (http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html#BSS)