Using Ensemble Models for Probabilistic Turbulence Forecasting at the Aviation Weather Center/Aviation Weather Testbed

Brian P. Pettegrew[1] and David Bright[2]

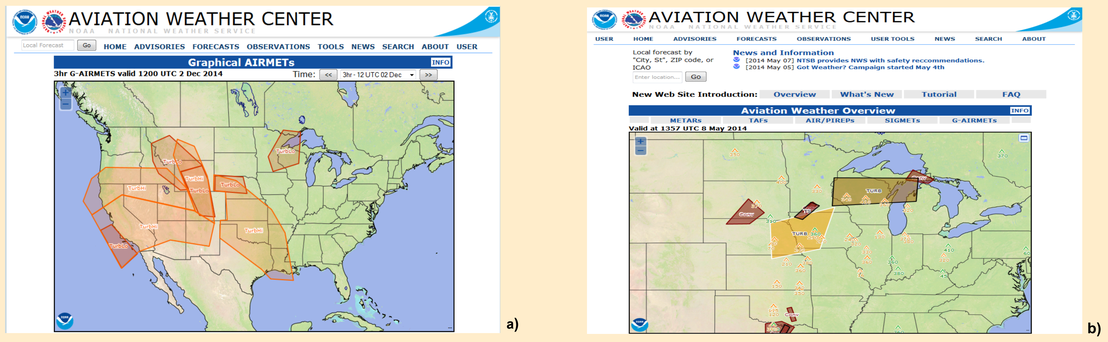
[1]Cooperative Institute for Research in the Atmosphere, Ft. Collins, Colorado, USA
[2]NOAA/NWS/NCEP Aviation Weather Center, Kansas City, Missouri, USA

Introduction

The National Weather Service's Aviation Weather Center (AWC) is tasked with the production of G-AIRMETs to identify flight hazards such as turbulence, icing, and instrument flight rules, in order to maximizing the safety and efficiency of flight. While many independent turbulence diagnostics exist (e.g. Ellrod-Knapp, Ellrod-Knox, TKE, etc.), the emphasis in recent years has been to create a multi-diagnostic approach for a more complete solution (e.g. Sharman et al., 2006). Instead of utilizing a deterministic model, a multi-diagnostic, probabilistic forecast has been constructed using the Short Range Ensemble Forecast (SREF) system. The end product is based on the frequency of member agreement among the individual SREF members. Diagnostic weights and calculations vary based on the vertical level and extends from near the surface to the stratosphere. This algorithm has been input into the Aviation Weather Testbed Ensemble Post-Processor for experimental use, evaluation, and training of operational forecasters at the AWC. Initial verification, both subjectively from forecaster use, and objectively show promise.

Operational Products

The Aviation Weather Center (AWC) currently produces official forecasts for all major aviation hazard, including turbulence, icing, ceiling and visibility and convection. Considering only turbulence, the AWC issues 3-hourly snapshots of moderate-or-greater Clear Air Turbulence (CAT) in the form of Graphical AIRMETs. SIGMETs are issued as needed for areas of severe-or-greater CAT (Figure 1).



a) b)

Figure 1. a) GAIRMETs for high and low-level (above and below 18000 feet, respectively) clear air turbulence (CAT) over the conterminous U.S. (CONUS) as issued by the AWC. **b)** SIGMETs for CAT overlaid with PIREPs as issued by the AWC.

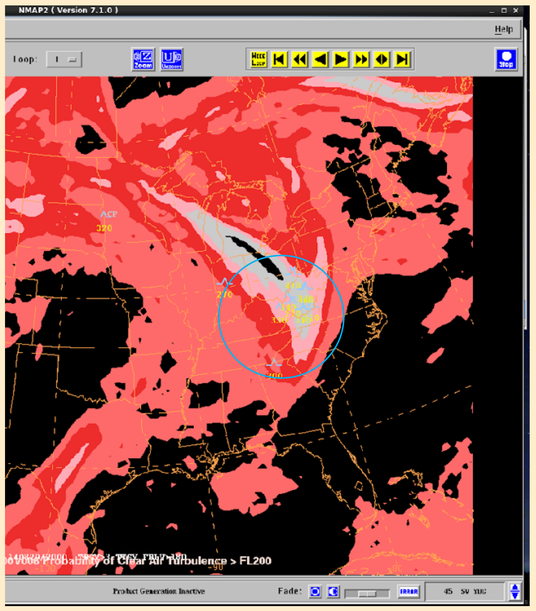


Figure 2. Probability of CAT above 20000 ft, 6 hour forecast valid at 2100 UTC 20 August 2014 with moderate-or-greater PIREP observations valid from 2000 and 2100 UTC 20 August 2014.

Objectives

The objective of developing a probabilistic view of CAT at the AWC are twofold. The first is improving forecasting and decision support by meteorologists at the AWC. Many forecasters still rely on a single deterministic solution of Ellrod or Ellrod-Knox in addition to experience and pattern recognition. The second objective is to address the report by the National Transportation and Safety Board (NTSB) to the National Weather Service calling for increased consistency in NWS aviation products and services. The use of ensemble-based forecasting leads to a more consistent forecast and forecast process.

Description

The AWC multi-diagnostic ensemble guidance utilizes the operational SREF from NCEP. Multiple turbulence diagnostics are calculated at each isobaric level on the SREF 40-km post-processed grid. Twenty-one diagnostics are used for upper level turbulence (above FL180) and eight diagnostics are used for lower level turbulence (below FL180). The diagnostics are assembled among all ensemble members and the result is a relative frequency (hereafter referred to as a "probability"). A threshold is applied to each diagnostic across each SREF member, with the probability derived from the fraction of diagnostic and member agreement. The thresholds used to diagnose the probabilities for each turbulence calculation are statistically determined. The guidance is then made available to forecasters in real-time through the Aviation Weather Testbed (AWT). An example is shown in Figure 2. It is also available online through the testbed website at http://testbed.aviationweather.gov.

Corresponding Author Address:
Brian P. Pettegrew
Cooperative Institute for Research in the Atmosphere
NOAA/NWS/NCEP Aviation Weather Center
7220 NW 101st Terrace
Kansas City, MO 64153
brian.p.pettegrew@noaa.gov

Disclaimer: The views expressed are those of the authors and do not necessarily represent the official policy and position of the U.S. Government.

Diagnostic Verification and Probability Thresholds

To determine a best threshold for constructing a probability, a statistical verification of probabilities constructed by each individual diagnostic was performed using PIREPs and EDR observations to determine a minimum Brier score using a range of thresholds. Deterministic values were calculated for each turbulence diagnostic for each SREF member and then a probability of that diagnostic was determined. For example, the Ellrod index was calculated for each SREF member and then a probability of Ellrod exceeding a threshold was determined. Probabilities were determined for a series of differing thresholds. The optimal threshold applied to the multi-diagnostic calculation was determined where the Brier score was minimized. In some cases, the score dropped quickly and then continued a slow decent in the range of chosen thresholds (see PVU diagnostic in Figure 3b below). In this case, the threshold at the base of quickest descent was chosen.

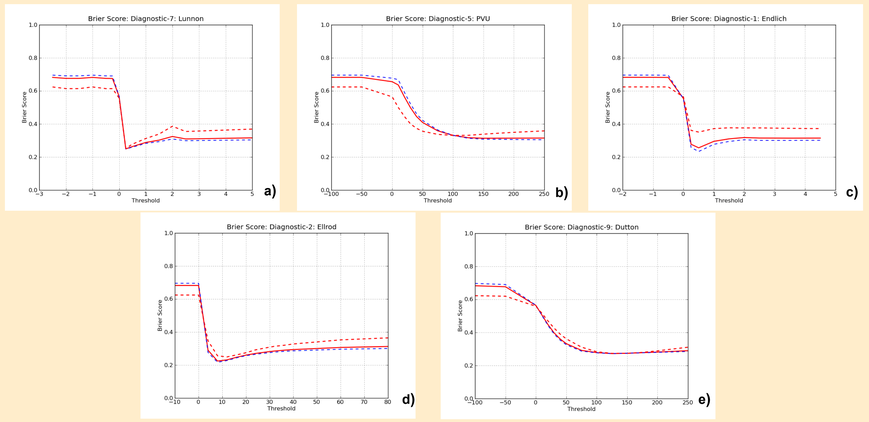


Figure 3. Brier scores for various diagnostics forecasting CAT: **a)** Lunnon, **b)** PVU, **c)** Endlich, **d)** Ellrod, and **e)** Dutton. Thresholds applied to each diagnostic are along the x axis and the calculated Brier Score along the y. The thresholds for each diagnostic are specific to the diagnostic itself (different diagnostics will have different thresholds due to differing units of measurement). The red dashed line represents Brier scores against PIREPs, the blue dashed line represents Brier scores against EDR observations, and the red solid line is Brier scores using combined PIREP and EDR.

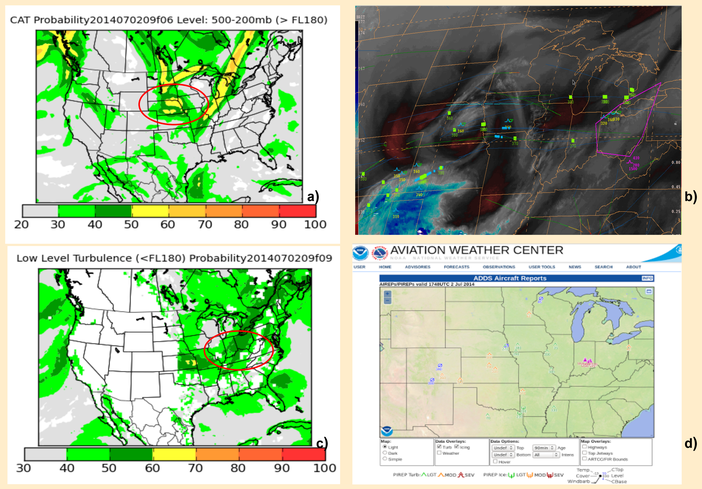


Figure 4. a) AWC Multi-diagnostic, ensemble probability of CAT above FL180 valid at 1500 UTC 2 July 2014 and **b)** water vapor image and G-AIRMET CAT forecasts valid at same with moderate-or-greater observations of turbulence (PIREPs and EDR). **c)** Probability of CAT below FL180 valid at 1800 UTC 2 July 2014 and **d)** moderate-or-greater observations of turbulence valid at same.

Summary and Evaluation

- Data collection for statistical verification and calibration continues.
  - Initial results show good resolution between probability bins
  - Distributions indicate greater amount of higher intensity EDR being captured by higher probabilities (Figure 5).
- Ensemble forecasts available for forecasters in AWT in real-time.
- Usage of ensemble forecasts increasing at AWC and the AWT, and construction of ensemble-based support tools for AWC's Domestic and International forecasts, advisories, and warnings will continue.

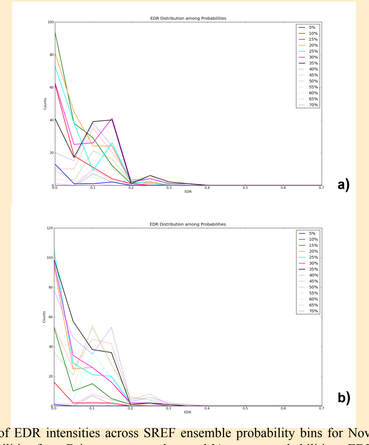


Figure 5. A distribution of EDR intensities across SREF ensemble probability bins for November 2014 **a)** threshold optimized probabilities from Brier score procedure and **b)** current probabilities. EDR intensities are along the x-axis while raw counts of observations are along the y-axis. While it is unclear what the moderate threshold of EDR is, a value of 0.15 and above is considered to be significant operationally.

Citation: 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.