

DESCRIPTION AND EVALUATION OF THE SECOND GENERATION GRAPHICAL TURBULENCE GUIDANCE FORECASTING SYSTEM

R. Sharman*, J. Wolff, and G. Wiener,
Research Applications Program
National Center for Atmospheric Research,
Boulder, CO

ABSTRACT

The first generation Graphical Turbulence Guidance (GTG1) became "operational" in March 2003 for use by meteorologists and airline dispatchers as an aid to forecasting areas of upper level (>FL200) clear-air turbulence (CAT) over the continental U.S. GTG uses a combination of individual turbulence diagnostics derived from NCEP's Rapid Update Cycle (RUC) forecasts weighted by comparisons to turbulence pilot reports. This technique of combining diagnostics has been shown to provide superior performance to the use of individual diagnostics alone in producing automated turbulence forecasts.

Since the original GTG release, development of the second generation GTG, GTG2, has been ongoing under sponsorship of the FAA Aviation Weather Research Program (AWRP). GTG2 includes several improvements over GTG1, including the use of new turbulence diagnostic algorithms, better thresholding of the individual diagnostics, and extension of forecasts to mid-levels (FL100-FL200). GTG2 is scheduled to become "operational" in early FY06. A brief description of the GTG2 product as well as preliminary statistical evaluations of GTG2 probabilities of detection performance are provided.

1. Introduction

Commercial (Part 121/129), air taxi (Part 135), and general aviation (GA - Part 91) encounters with turbulence continue to be a source of occupant injuries, and in the case of GA, often of fatalities and loss of aircraft. According to a recent MCR Federal survey of NTSB accident data for the years 1983-1997 (Eichenbaum, 1999), turbulence contributed to 664 accidents leading to 609 fatalities

(mostly GA), 239 serious and 584 minor injuries, for an estimated average annual societal cost of \$134 M. Although fatalities related to commercial airline turbulence encounters are almost nil (only one in this time period), turbulence encounters do account for a significant fraction (about 30%) of all weather related Part 121/129 incidents. The average number of air carrier turbulence-related injuries according to the NTSB records is about 45 per year, but these are of course only those that are reported to the NTSB. The actual number is probably higher: one major carrier reported almost 400 turbulence encounters over a 3 year period; another estimated about 200 turbulence related customer injury claims per year. Of significance to the subject at hand is that the MCR Federal report also estimated that *only about 30% of these upper level incidents were forecast based on previous turbulence pilot reports (PIREPs) or valid AIRMETs*. Clearly, even with this most liberal definition of a forecast, there is much room for improvement.

A large percentage of these turbulence encounters might be avoided if better turbulence forecast products were available for use by air traffic controllers, airline flight dispatchers, and flight crews. Strategic planning for turbulence avoidance can be accomplished if sufficiently accurate forecasts of turbulence are available. Previous studies (e.g., Fahey 1993) have shown that at least for commercial air carriers, strategic planning can reduce cabin injuries and reduce costs. However, current forecasting methods have not generally provided acceptably high detection rates and at the same time acceptably low false alarm rates. The term "acceptable" does not have a universal quantitative definition, but the Turbulence Joint Safety Implementation Team (JSIT) report to improve the quality of turbulence information, recommended probabilities of MOG detection should be > 0.8 and

* Corresponding author address:
sharman@ucar.edu

probabilities of null detection should be > 0.85 for turbulence forecasts to be most useful.

The lack of progress in the turbulence forecasting area is due in large part to the fact that, from the meteorological perspective, turbulence is a “microscale” phenomenon. In the atmosphere, turbulent “eddies” are contained in a spectrum of sizes, from 100s of kilometers down to centimeters. But aircraft bumpiness is most pronounced when the size of the turbulent eddies encountered are about the size of the aircraft; for commercial aircraft this would be eddy dimensions of about 100m or so. It is impossible to directly forecast atmospheric motion at this scale (it would take about 10m resolution with a grid-based forecast model), now or even in the foreseeable future. Fortunately, it appears that most of the energy associated with eddies of this scale cascades down from the larger scales of atmospheric motion (e.g. Dutton and Panofsky, 1970), which may in fact be resolved by current weather observations and numerical forecast models. Assuming the large-scale forecasts are sufficiently accurate, the turbulence forecasting problem is then one of identifying large-scale features that are conducive to the formation of aircraft scale eddies. So one major area of research over the last 50 years or so has involved efforts to establish a linkage between large-scale atmospheric features (i.e., observable by routine meteorological observations and resolvable by numerical weather prediction models) and aircraft-scale turbulence. Some of these linkages have been inferred through the efforts of National Weather Service and airline meteorological forecasters as turbulence forecasting rules-of-thumb, but the skill depends on the forecaster, and that skill diminishes rapidly with forecast lead time. Because there is now a tremendous amount of meteorological data available to forecasters, more than can be digested in a reasonable length of time, automated turbulence forecast tools are being developed to aid the human in making decisions about where to locate regions of potential turbulence that may be hazardous to aircraft.

To address the need for automated turbulence forecasting tools NCAR/RAP and NOAA/FSL, under sponsorship from the FAA’s Aviation Weather Research Program (AWRP), have been developing and testing a completely automated turbulence forecasting

system. This system was originally named the Integrated Turbulence Forecasting Algorithm, ITFA, and concentrated only on upper level ($>FL200$) clear-air turbulence (see Sharman and Comman, 1998, Sharman et al., 1999, Sharman et al., 2000, Sharman et al., 2002, Tebaldi et al., 2002). Mountain wave turbulence and convective sources of turbulence are not explicitly accounted for. The ITFA system became “operational” for qualified meteorologists and dispatchers in March 2003 and at that time was renamed the Graphical Turbulence Guidance (GTG) product. The first generation GTG, or GTG1, provides gridded CAT forecasts stratified by flight levels and graphical displays of turbulence potential are provided on those flight levels to interested users on the Aviation Digital Data Service (ADDS) website: (<http://adds.aviationweather.gov/turbulence>).

GTG2 expands the capabilities of GTG1 by extending turbulence analyses and forecasts down to FL100 from the GTG1 value of FL200. The FL100-FL200 altitude band is especially significant for air taxi commercial carriers. Thus in the new system, there are turbulence predictions at both upper levels ($\geq FL200$) and mid-levels (FL100-FL200). In addition, some new turbulence diagnostics were included as a result of continued turbulence diagnostic research. Within GTG2, the mid-level and upper-level forecasts are computed separately, and the results merged at the FL200 boundary. Within each major altitude band, the technique used to derive the forecasts is pretty much the same as that used to produce the GTG1 CAT forecasts. This technique has been described in the references cited above, but for the sake of completeness, will be reviewed in Section 2. Preliminary results of verification studies based on 0 and 6-hour RUC forecasts over the winter 2002-2003 season are presented in Section 3. Development and tuning is an ongoing task, and current problems and work areas are detailed in Section 4.

2. GTG Procedure

The GTG process is shown schematically in Fig. 1. The process starts by automatically ingesting gridded NWP data, which is supposed to accurately represent the large scale features of the atmosphere that may be

related to aircraft-scale turbulence. In principle, any NWP model could be used, but NCEP's Rapid Update Cycle (RUC) model was chosen because of the higher effective vertical resolution provided by the potential temperature ("theta") vertical coordinate system at upper levels in the model (Benjamin et al., 2004). The body of the procedure is in the box marked "GTG nowcast and forecast generator", which produces the GTG forecasts tuned to available observations in the form of turbulence pilot reports (PIREPs).

The essence of the GTG forecasting method is to use a combination of many separate turbulence diagnostics, with each diagnostic weighted to get best agreement with available observations (i.e., PIREPs). This idea of using a weighted combination of diagnostics to provide turbulence forecasts is not in itself a new one. For example, Dutton (1980) evaluated the performance of 11 diagnostics compared to pilot reports of CAT over the North Atlantic and parts of Europe. He found the weighted sum of the vertical and horizontal wind shears provided the best agreement with his observations. Also Clark et al. (1975) used a set of 5 weighted diagnostics, where the set used depended on the elevation band, and with the weights determined by the best fit of XB-70 stratospheric turbulence encounters over the western U.S. Similar procedures have been used by Russian investigators. For example, Leshkevich (1988) used a weighted sum of 12 diagnostics, and Buldovskii et al. (1976) used a weighted combination of horizontal temperature gradient and vertical wind shear to predict CAT, again with the weights determined by best agreement to available observations. However, all of these studies were based on a limited set of observations and the weights determined by the best fit to this limited set. These weights, once established, are static, i.e., they never change. The GTG also obtains weights for a set of diagnostics based on the best fit to observations, but these weights are dynamically determined and updated every 3 hours to give the best agreement with the current set of available PIREPs. This approach of dynamically assigning weights has been shown by Tebaldi et al. (2002) to give better performance than using a single diagnostic.

The entire GTG process involves the following six step procedure. Each step is

modular, both in function and software implementation, and each step has various options which users may invoke.

Step 1. Compute a set of n turbulence diagnostics D_n (e.g. a local Richardson number) from NWP output at each grid point in the model domain at assimilation time. Currently, GTG2 uses a combination of 10 turbulence diagnostics at both upper and mid-levels. The suite of diagnostics chosen depends on the overall performance of each diagnostic separately and also on the desire to ensure the diagnostics used are actually attempting to identify different atmospheric processes that may be contributing to turbulence, i.e., the diagnostics are uncorrelated with one another. Some experimentation is still in progress to determine the optimal set of indices, but currently, at upper levels, the 10 used are:

1. Colson-Panofsky index (Colson and Panofsky 1965)
2. Richardson number (e.g., Endlich 1964, Kronebach 1964, Dutton and Panofsky, 1970)
3. DTF3 (Marroquin 1998)
4. 2D frontogenesis function (isentropic coordinates) (e.g., Bluestein, 1992, vol. 2)
5. Unbalanced flow diagnostic (Knox, 1997; McCann 2001; Koch and Caracena 2002)
6. Horizontal temperature gradient (Buldovskij et. al. 1976)
7. Stone (Stone 1966, Knox 1997)
8. NCSUi (Kaplan et al. 2004)
9. Horizontal velocity eddy dissipation rate (Frehlich and Sharman 2004, Lindborg 1999)
10. Vertical velocity eddy dissipation rate (Frehlich and Sharman 2004, Lindborg 1999)

And at mid-levels the 10 used are:

1. Ellrod (Ellrod and Knapp 1992)
2. DTF3 (Marroquin 1998)
3. MOSS predictor (Reap 1996)
4. ABSIA (McCann 2001)
5. Horizontal temperature gradient (Buldovskij et. al. 1976)
6. Wind speed (Endlich 1964)
7. NCSUi (Kaplan et al. 2004)
8. Horizontal velocity eddy dissipation rate (Frehlich and Sharman 2004, Lindborg 1999)

9. Vertical velocity eddy dissipation rate (Frehlich and Sharman 2004, Lindborg 1999)
10. 2D frontogenesis function (pressure coordinates) (e.g., Bluestein, 1992, vol. 2)

Step 2. Map the D_n to a common turbulence intensity scale D_n^* such that all have the same range, 0-1, where 0 predicts no turbulence (null), 1 predicts extreme turbulence, and 0.5 predicts moderate turbulence. The same scale is also used for PIREP intensity to allow intensity comparisons. A required input for combining the various turbulence diagnostics is the threshold values that distinguish the null-light, light-moderate, moderate-severe, and severe-extreme turbulence categories. These thresholds are derived by comparing PIREP values to index values for many index-PIREP comparisons, and taking the median of the index value corresponding to each PIREP turbulence category.

Step 3. Compare each diagnostic to the available observations (PIREPs) within a time window (currently ± 90 min.) around the NWP model assimilation time for the altitude band of interest by forming a “score” for each diagnostic that measures the relative error between the intensity as computed by each diagnostic and all available PIREPs. For example, if diagnostic *A* near the location of a moderate intensity PIREP was above its threshold for moderate turbulence but below its threshold for severe turbulence the score assigned to *A* would be relatively high, whereas if diagnostic *B* was below the light intensity threshold at the same PIREP location, the score assigned would be relatively low.

Step 4. Form a set of weights W_n for each diagnostic *n* proportional to the score derived from step 3. Because the number of PIREPs available at any time is still a small number (usually around 50 or so), it is not possible to form weights regionally or vertically, so the weights assigned are constant throughout the domain of interest.

Step 5. Combine the weighted diagnostics to form the GTG combination.

$$GTG = W_1 D_1^* + W_2 D_2^* + W_3 D_3^* + \dots + W_n D_n^*$$

An example is shown schematically in Figure 2.

Step 6. Use these weights in conjunction with the NWP forecast data (RUC provides 3,6,9,12 hr forecasts) to derive GTG forecasts.

Currently, this entire cycle repeats itself with every major NWP update; for RUC this is every 3 hours.

Within GTG2, mid-level and upper-level forecasts are computed separately, and the results merged at the FL200 boundary. This was necessary since it was found that (1) the best set (in terms of PODY-PODN performance) of turbulence diagnostics was not the same at upper- and mid-levels, (2) their optimum threshold values were not the same either, (3) the number of available PIREPs was substantially less at mid-levels than at upper levels, thus requiring different PIREP time windows to be used in the two altitude regimes.

3. Prototype GTG2 performance statistics

The accuracy of derived values for the GTG combination and the individual diagnostics within GTG can be assessed from the only routine observations of aircraft scale atmospheric turbulence available, reports of encounters with turbulence by commercial airline pilots. Pilot reports are semi-automated and give information about a turbulence encounter (time, latitude, longitude, altitude, severity). There is some subjectivity associated with these reports, especially with regard to severity, and it must be realized that the report is based on a turbulence experience along a flight path, i.e. along a line, and is usually reported as a single point value. If the model-derived diagnostics are supposed to be a grid point average, the correspondence to a line is not necessary direct. Nevertheless, the relative performance of various diagnostics can be evaluated by comparisons to turbulence pilot reports as in Tebaldi et al. (2002). In that study the metric used to evaluate the performance of various turbulence diagnostics was the area contained under probability of detection (POD) curves, similar to radar operating characteristic curves. In this procedure a set of thresholds is assumed for each diagnostic, and given that threshold, the diagnostic

performance based on comparisons to available turbulence pilot reports is evaluated for both null (as measured by PODN, the fraction of null events correctly detected) and moderate or greater turbulence reports (as measured by PODY, the fraction of moderate or greater turbulence events correctly detected). For small values of the chosen threshold, PODY will obviously be high, near unity, while PODN will be low, near 0, and vice versa for large values of the chosen threshold. For the range of thresholds selected, higher combinations of PODY and PODN and therefore larger areas under the PODY-PODN curves, imply greater skill in discriminating between null and moderate-or-greater turbulence events.

A sample of the PODY-N statistical performance derived during initial testing of GTG2 is provided in Figure 3 (upper levels) and Figure 4 (mid levels). The curves are based on 3 months of RUC20 input data for the period 1 Nov 2002-31 Jan 2003, for both 0-hr and 6-hr forecasts. As expected, the GTG combination is superior to the individual diagnostics by this performance metric for all cases. For both forecast times (0-hr and 6-hr) the upper level performance is slightly better than the mid-level performance. This is probably due in part to the fact that there are fewer PIREPs available to fit the data at mid-levels compared to upper-levels and in part due to the fact that experience derived from the GTG1 development has allowed us to formulate better turbulence diagnostics at upper levels. In fact, in Fig. 4 for mid-levels, the curves are not as smooth as the upper level curves due to the fewer number of PIREPs available for assessment. Because of the small number of PIREPS available at mid levels two regression strategies were tested: one using the best set of weights obtainable at analysis time (red curves in Fig. 4), and another using a constant set of weights based on the overall performance of the individual diagnostics (blue curves in Fig. 4). Given the small difference in forecast performance of these two methods the simpler method of combining indices will be used in further test and development.

Of course the day-to-day forecast performance of the GTG combination will vary with the performance of the forecast diagnostics. Figure 5 shows a time series of upper-level 6-hr forecast True Skill Score (TSS) performance for the GTG2 combination

(in red) and the individual diagnostics (in black) for a chosen set of thresholds over the 90 days of the test period. The performance of all diagnostics is highly variable, although the GTG combination is usually somewhat better than any individual diagnostic, and sometimes much better.

4. Discussion

The ability to provide accurate aircraft scale turbulence nowcasts and forecasts is hampered by several fundamental difficulties. First, the resolution of current NWP models (several 10s to 100 km roughly) is about two orders of magnitude too coarse to resolve aircraft scale turbulence (roughly 100s m). Therefore, aircraft scale turbulence diagnoses and predictions must be based on resolvable scale features. However, and this is the second difficulty, the performance of turbulence diagnostics is hampered by our current lack of understanding of the linkage between NWP observable scale features and aircraft scale turbulence. An implicit assumption in all these diagnostics is that turbulence generating mechanisms have their origin at resolvable scales and the energy cascades down to aircraft scales, but it is unclear what the exact mechanism is that creates small scale motion from the larger scales. Third, even if it is true that aircraft scale turbulence has its origins at the resolvable scales, the turbulence forecast system has all the inherent NWP errors associated with the resolvable scales. Fourth, it is not clear that the current suite of turbulence diagnostics is in fact capturing all the relevant information that the larger scale representations can provide. Finally, there is the difficult matter of verification. In the GTG system we are using PIREPs for tuning and verification. But an individual PIREP is subject to spatial and temporal errors, and is subjective in its intensity rating. Further, the PIREPs are variable in space and time, and in particular undergo a strong diurnal period (considerably fewer at night) making it difficult to perform consistent tuning and verifications over all time periods. The quantitative automated in-situ turbulence reporting system (Cornman, et al., 1995) should eliminate most of the uncertainty associated with PIREPs but will still not fill in the gaps at night.

Nevertheless, the overall performance of the GTG combination, although not optimal, is

still skillful enough to provide useful information to meteorologists and dispatchers for strategic planning purposes. At the same time there continues to be concentrated efforts to provide a better turbulence forecasting system through the following research areas:

- Better diagnostics. This is a continued research area in the major laboratories and universities. But any diagnostic must be judged by its overall performance, not just on a few selected cases. In addition, information about when a particular diagnostic performs well and when it does not, could be used in dynamically assigning its weight within the GTG framework. But this situational dependence can only be assessed through careful case studies which tend to be human intensive. Also, diagnostics for other sources of turbulence, e.g., mountain wave induced turbulence or convectively induced turbulence need to be developed and tested.
- “Local” fits. Within the current GTG framework, the best fit of diagnostics is determined for the entire volume of atmosphere between the altitude bands of interest. Better fits are probably attainable to subvolumes, which could be overlapped to give smooth transitions from one subvolume to another. Although the number of PIREPs available for regional or local fits is probably insufficient at the current time, the use of the turbulence in-situ measurements (with one-minute sampling intervals in cruise) may allow local fits, both horizontally and vertically.
- Better optimization strategies. Although several optimization or weighting strategies have been tried (e.g., Sharman et al. 2002, Tebaldi et al. 2002), there are many others available and it may be the one of these methods leads to demonstrably better performance. Also, better methods may be derived for combining indices when several sets of indices are intended to describe one turbulence generation source, and another set describes a different generation source.

Acknowledgments. This research is 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.

REFERENCES

- 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.
- Bluestein, H. B., 1992: Synoptic-Dynamic Meteorology in Midlatitudes, Vols. I and II. Oxford Univ. Press.
- Buldovskii, G. S., S. A. Bortnikov, and M. V. Rubinstejn, 1976: Forecasting zones of intense turbulence in the upper troposphere. *Meteorologiya i Gidrologiya*, **2**, 9-18.
- Clark, T. L., J. R. Scoggins, and R. E. Cox, 1975: Distinguishing between CAT and non-CAT areas by use of discriminant functional analysis. *Mon. Wea. Rev.*, **103**, 514-520.
- Cornman, L. B., C. S. Morse, and G. Cuning, 1995: Real-time estimation of atmospheric turbulence severity from in-situ aircraft measurements. *J. Aircraft*, **32**, 171-177.
- Dutton, M. J. O., 1980: Probability forecasts of clear-air turbulence based on numerical output. *Meteor. Mag*, **109**, 293-310.
- Dutton, J. and H. A. Panofsky, 1970: Clear Air Turbulence: A mystery may be unfolding. *Science*, **167**, 937-944.
- Eichenbaum, H., 2000: Historical overview of turbulence accidents. MCR Federal, Inc. report TR-7100/023-1.
- Ellrod, G. P. and D. L. Knapp, 1992: An objective clear-air turbulence forecasting technique: verification and operational use. *Wea. Forecasting*, **7**, 150-165.
- Endlich, R. M., 1964: The mesoscale structure of some regions of clear-air turbulence. *J. Appl. Meteor.*, **3**, 261-276.
- Fahey, T. H., 1993: Northwest Airlines atmospheric hazards advisory & avoidance system. Preprints, *Fifth Conf. on Aviation Weather Systems*, 409-413, American Meteorological Society, Vienna, VA.
- Frehlich, R. and R. Sharman, 2004: Estimates of turbulence from numerical weather prediction model output with

applications to turbulence diagnosis and data assimilation. *Mon. Wea. Rev.*, in press.

Kaplan, M. L., K. M. Lux, J. D. Cetola, A. W. Huffman, J. J. Charney, A. J. Riordan, S. W. Slusser, Y.-L. Lin, and K. T. Waight III, 2004: Characterizing the severe turbulence environments associated with commercial aviation accidents. A Real-Time Turbulence Model (RTTM) designed for the operational prediction of hazardous aviation turbulence environments. Submitted to *Meteor. Atmos. Phys.*

Koch, S. E. and F. Caracena, 2002: Predicting clear-air turbulence from diagnosis of unbalance flow. Preprints, *Tenth Conf. on Aviation, Range, and Aerospace Meteorology*, 359-363. American Meteorological Society, Portland, OR.

Knox, J. A. 1997: Generalized nonlinear balance criteria and inertial stability. *Mon. Wea. Rev.*, **54**, 967-985.

Leshkevich, T. V., 1988: Automated method of predicting the probability of clear-air turbulence. *Meteorologiya i Gidrologiya*, **10**, 27-33.

Lindborg, E., 1999: Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence? *J. Fluid Mech.*, **388**, 259-288.

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

Lindborg, E., 1999: Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence? *J. Fluid Mech.*, **388**, 259-288.

Marroquin, A., 1998: An advanced algorithm to diagnose atmospheric turbulence using numerical model output. Preprints, *Sixteenth Conf. on Weather Analysis and Forecasting*, 79-81. Phoenix, AZ. American Meteorological Society.

McCann, D. W., 2001: Gravity waves, unbalanced flow, and aircraft clear air turbulence. *National Weather Digest*, **25**, 3-14.

Reap, R. M., 1996: Probability forecasts of clear-air turbulence for the contiguous U.S. *National Weather Service Office of Meteorology Technical Procedures Bulletin No. 430*, NOAA, 15 pp.

Sharman, R. and L. Cornman, 1998: An integrated approach to clear-air turbulence prediction. **AIAA 98-0382**. *AIAA 36th Aerospace Sciences Meeting & Exhibit, 10-13 January 1998*, AIAA, Reno, NV.

Sharman, R., C. Tebaldi, and B. Brown, 1999: An integrated approach to clear-air turbulence forecasting. Preprints, *Eighth Conf. on Aviation, Range, and Aerospace Meteorology*, 68-71. American Meteorological Society, Dallas, TX.

Sharman, R., G. Wiener, and B. Brown, 2000: Description and verification of the NCAR Integrated Turbulence Forecasting Algorithm (ITFA). **AIAA 00-0493**. *AIAA 38th Aerospace Sciences Meeting & Exhibit, 10-13 January 2000*, Reno, NV.

Sharman, R., C. Tebaldi, J. Wolff, G. Wiener, 2002: Results from the NCAR Integrated Turbulence Forecasting Algorithm (ITFA) for predicting upper-level clear-air turbulence. Preprints, *Tenth Conf. on Aviation, Range, and Aerospace Meteorology*, 351-354. American Meteorological Society, Portland, OR

Stone, P. H., 1966: On non-geostrophic baroclinic stability. *J. Atmos. Sci.*, **23**, 390-400.

Tebaldi, C., D. Nychka, B. G. Brown, and R. Sharman, 2002: Flexible discriminant techniques for forecasting clear-air turbulence. *Environmetrics 2002*, **13**, 859-878.

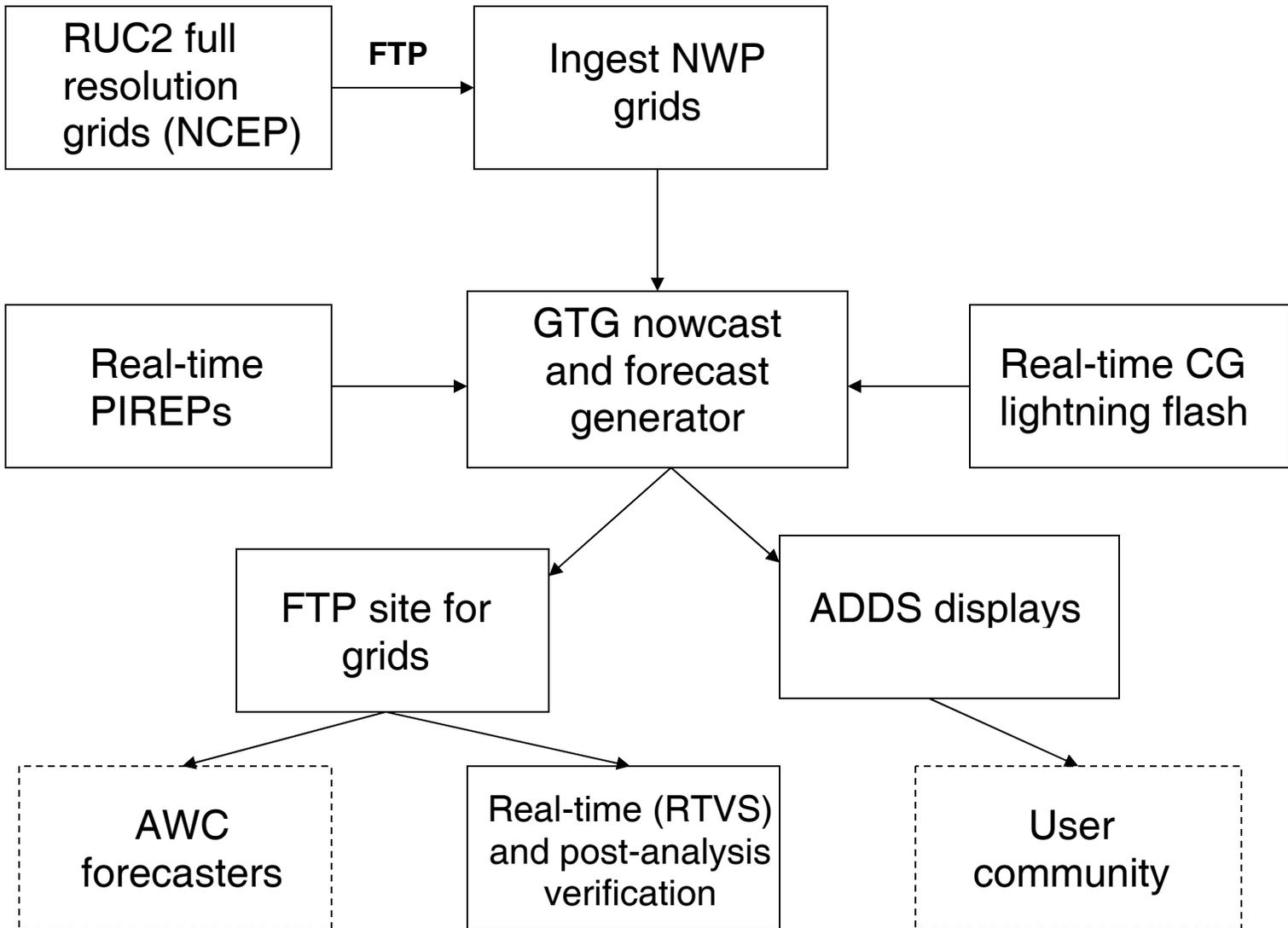


Figure 1. GTG process, including inputs and outputs.

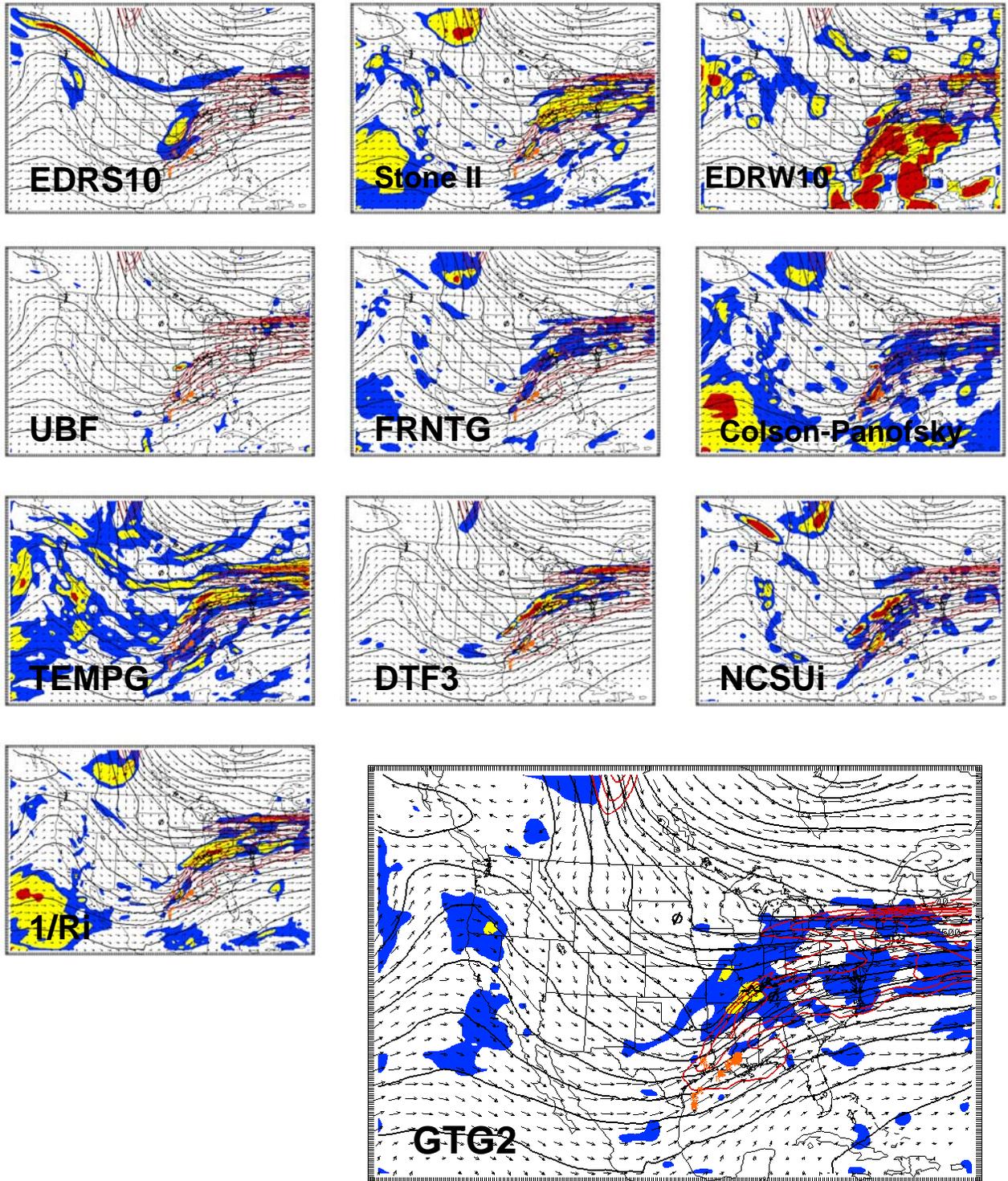


Figure 2. Contours (color filled) of turbulence predicted intensity for 10 different indices and the GTG2 composite (lower right) for a 6 hour forecast at FL250 valid 0Z 5 Nov 2002. For reference, PIREPs, wind vectors and height lines are superposed on each figure.

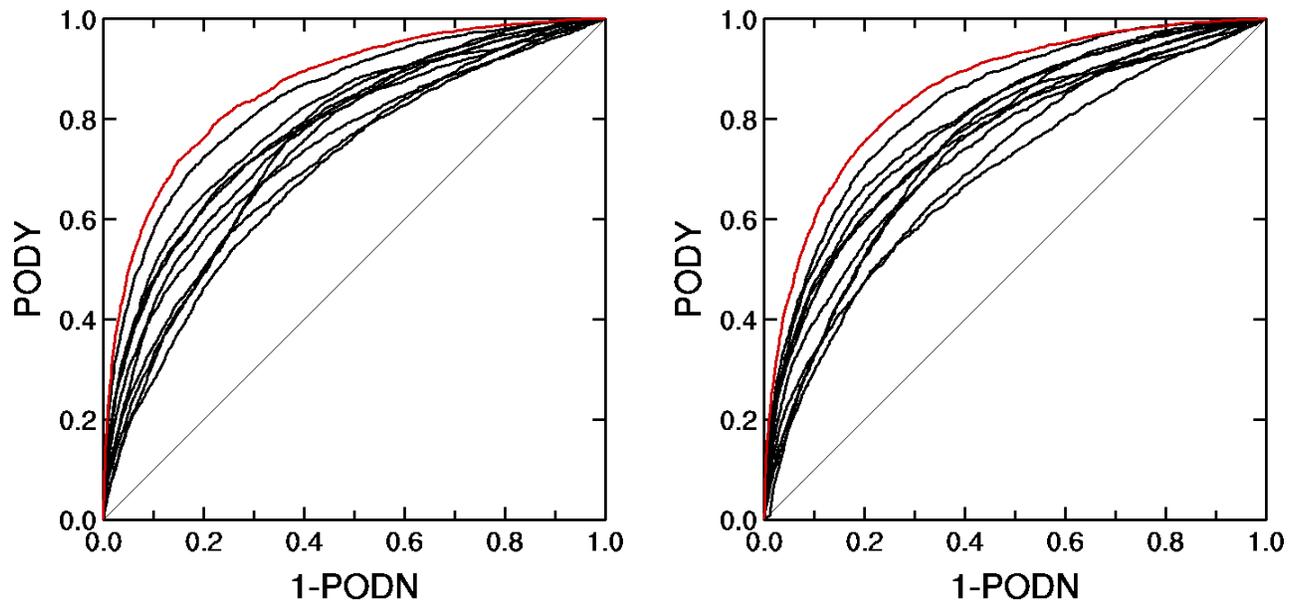


Figure 3. Comparison of the prototype GTG2 PODY-PODN performance statistics (individual diagnostics in black, GTG combination in red) derived from 3 months of 18Z analyses (0-hour forecasts) (left panel) and 18Z 6-hour forecasts (valid 0Z) (right panel), for upper levels (>FL200).

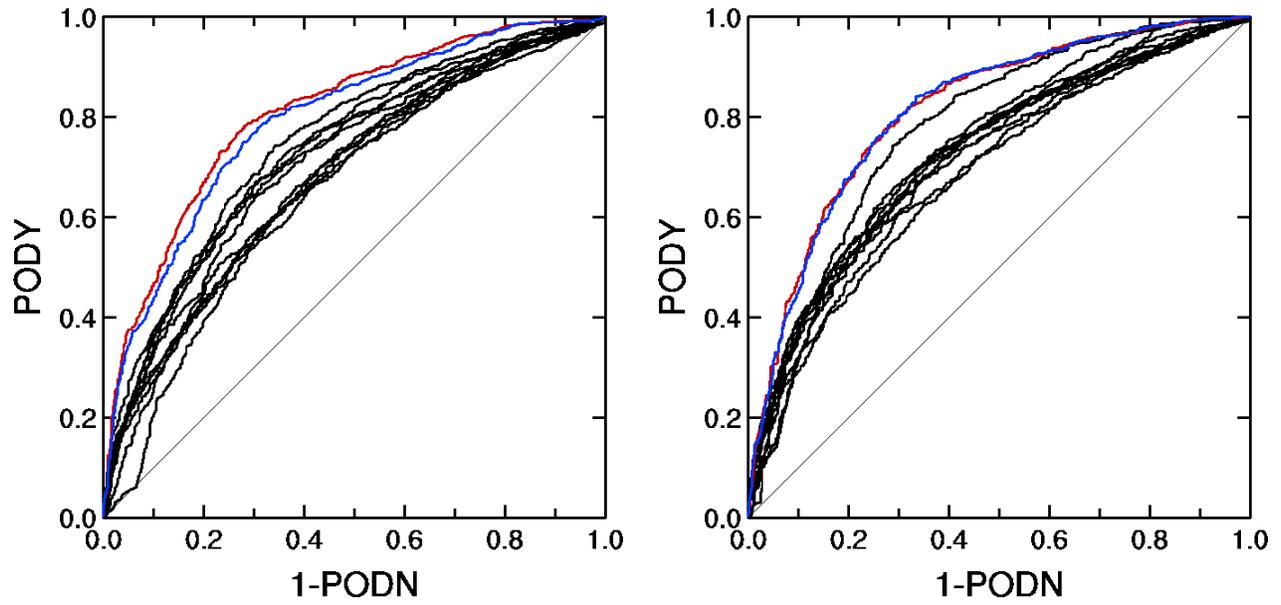


Figure 4. Comparison of the prototype GTG2 PODY-PODN performance statistics (individual diagnostics in black, GTG combination in blue, red) derived from 3 months of 18Z analyses (0-hour forecasts) (left panel) and 18Z 6-hour forecasts (valid 0Z) (right panel), for mid levels (FL100-FL200).

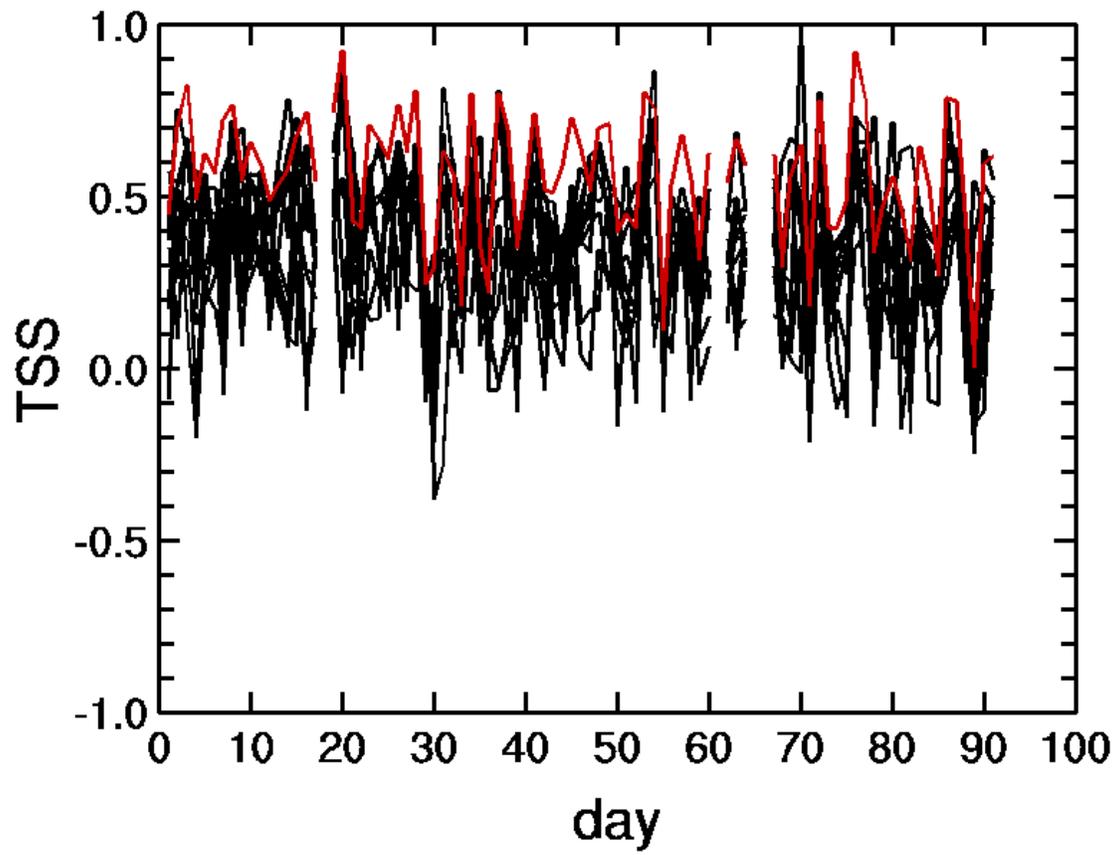


Figure 5. GTG (red) and individual diagnostics (black) daily TSS of 6-hr upper level forecasts.