NOAA Technical Memorandum OAR GSD-54

https://doi.org/10.7289/V5/TM-OAR-GSD-54

Assessment of the Graphical Turbulence Guidance, Nowcast (GTGN)

March 2016

Matthew S. Wandishin
Laura Paulik
Joan Hart
Brian Etherton
Melissa A. Petty

Earth System Research Laboratory
Global System Division
Boulder, Colorado
March 2016

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / Office of Oceanic and Atmospheric Research

https://doi.org/10.7289/V5/TM-OAR-GSD-54

NOAA Technical Memorandum OAR GSD-54

Assessment of the Graphical Turbulence Guidance, Nowcast (GTGN)

Matthew S. Wandishin[2]
Laura Paulik[2]
Joan Hart[2]
Brian Etherton[1]
Melissa A. Petty[3]

[1] National Oceanic and Atmospheric Administration, Earth System Research Laboratory, Global Systems Division (NOAA/ESRL/GSD)

[2] Cooperative Institute for Research in Environmental Sciences (CIRES) and NOAA/ESRL/GSD

[3] Cooperative Institute for Research in the Atmosphere (CIRA) and NOAA/ESRL/GSD

Acknowledgements

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. The authors would like to thank the Turbulence Product Development Team for providing the forecast data that was needed for the evaluation.

UNITED STATES DEPARTMENT OF COMMERCE

Wilbur Ross
Secretary

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

RDML Tim Gallaudet, Ph.D., USN Ret.
Under Secretary for Oceans
And Atmosphere/NOAA Administrator

Office of Oceanic and Atmospheric Research

Craig N. McLean
Assistant Administrator

EXECUTIVE SUMMARY

The QA PDT was tasked to assess the quality of the Graphical Turbulence Guidance, Nowcast, (GTGN) algorithm developed by the National Center for Atmospheric Research. This product is designed to provide near-real-time situational awareness to support operational aviation turbulence decisions. GTGN uses a short-term (1- or 2-h lead) forecast from the Graphical Turbulence Guidance, Version 3 (GTG3) product as a first guess field which is then augmented with direct observations—both pilot reports (PIREPs) and EDR measurements—and remotely sensed data from the NEXRAD Turbulence Detection Algorithm (NTDA). Updates are supplied every 15 minutes, with output provided on the same Rapid Refresh (RAP)-based, 13-km grid used by GTG3.

GTGN is intended as a nowcast product to be used for tactical decisions. GTG3 provides forecasts out to 18h to support strategic decision-making. In other words, these two products are meant to provide complementary information. There is, currently, no gridded product providing real-time situational awareness for atmospheric turbulence; the best available product is a short-term GTG3 forecast. Therefore, they are compared here as competing products.

The assessment compares GTGN with the 2-h GTG3 forecast used in its first-guess field and incorporates output from the operational GTG3 algorithm, GTGN, as well as PIREPs and EDR values derived from *in situ* measurements. The forecasts were analyzed using output generated from 1 July – 30 September 2013 and 1 January – 31 March 2014 over the CONUS. Primary findings include:

- When assessed in the context of near-real-time situational awareness, GTGN outperforms GTG3
- GTGN has more strong turbulence and more smooth turbulence than GTG3
- In winter, GTGN has fewer misses and fewer false alarms than GTG3
- GTGN recovers much of the decline in skill from winter to summer seen with GTG3 (by capturing events missed by GTG3), but a lower forecast threshold is required to achieve that skill
- Results vary somewhat by region and by altitude layer:
    - Improvement is greatest in the Southeast region
    - In winter, improvement is greatest in 0-10 kft layer
    - In summer, improvement is greatest in 10-20 kft layer
- Fall off in skill for GTGN is slow—45 min old GTGN still much better than the corresponding GTG3

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

The QA PDT was tasked to assess the quality of the Graphical Turbulence Guidance, Nowcast, (GTGN) algorithm developed by the National Center for Atmospheric Research (NCAR). This product is designed to provide near-real-time situational awareness to support operational aviation turbulence decisions. GTGN uses a short-term (1- or 2-h lead) forecast from the Graphical Turbulence Guidance, Version 3 (GTG3) product as a first-guess field which is then augmented with direct observations—both pilot reports (PIREPs) and Eddy Dissipation Rate (EDR) measurements—and remotely sensed data from the NEXRAD Turbulence Detection Algorithm (NTDA). Null and moderate-or-greater (MOG) turbulence observations are given more weight than reports of light (LGT) turbulence. Updates are supplied every 15 minutes, with output provided on the same Rapid Refresh (RAP)- based, 13-km grid used by GTG3.

The assessment compares GTGN with the 2-h GTG3 forecast used in its first-guess field and incorporates output from the operational GTG3 algorithm, GTGN, as well as PIREPs and EDR values derived from *in situ* measurements. The assessment consists of four main areas of investigation:

1. Overall comparison between GTGN and GTG3
2. Sensitivity of performance by season: summer vs. winter
3. Sensitivity of performance by altitude layer
4. Sensitivity of performance by geographic region

The results and conclusions obtained from the QA PDT assessment will be provided to a Technical Review Panel as input to the decision on whether the GTGN algorithm is ready for transition to operations at the National Weather Service (NWS).

2 DATA

This section describes the forecast and observation data that are included in the assessment, along with the principal stratifications that are used. The time period for this study consists of a winter period, 1 January – 31 March 2014, and a summer period, 1 July – 30 September of 2013.

2.1 FORECASTS

GTGN and GTG3 are produced on the same RAP-based grid with 13-km horizontal resolution and 1000-ft vertical resolution (with the lowest level at 100 ft rather than zero). GTG3 is run every hour with lead times out to 18 h, but only the 2-h lead will be used here. GTGN is produced every 15 minutes and is valid for the subsequent 15-minute period, with the [HH-1]30, [HH-1]45, HH00, and HH15 issuances all sharing the same HH00-valid GTG3 forecast.

2.2 OBSERVATIONS

2.2.1 PILOT REPORTS (PIREPS)

PIREPs are reported irregularly at the pilot's discretion and include a subjective assessment of many meteorological variables including the existence/absence of turbulence and a subjective measure of the turbulence intensity. Included in the turbulence reports are the location, altitude or range of altitudes, type of aircraft, air temperature, and intensity of turbulence (NWS 2007). Additionally, PIREPs include optional pilot remarks that are sometimes used to identify the source of the encountered turbulence, e.g., mountain waves.

2.2.2 *IN SITU* MEASUREMENTS

EDR is the International Civil Aviation Organization (ICAO) standard for automated reporting of turbulence from commercial aircraft. The values are derived from *in situ* measurements from a number of United Airlines (UAL) 737 and 757 and Delta Airlines (DAL) 737 and 767 aircraft. The derivation and reporting methods are different between the two airlines.

For the UAL aircraft, on-board equipment measures and reports vertical accelerations of the aircraft. These measurements are converted into an EDR value and then reported back to a database where they undergo quality control processes. The EDR observing system reports a maximum and median value every minute in 0.1-width bins. Due to equipment sensitivity during ascent/descent stages of flight, EDR observations below 20000 ft are not utilized (Cornman et al. 2004).

EDR values from DAL aircraft are computed directly from the vertical wind measurements. Reports consist of “heartbeat” reports issued every 15 minutes after takeoff, and “triggered” reports, issued whenever one of the following three conditions are met:

1. A single peak EDR value > 0.18
2. Three out of six peak EDR values > 0.12
3. Four out of six mean EDR values > 0.08

Triggered reports provide the previous six minutes of EDR values (at one-minute resolution), while reports triggered by either of the first two conditions also include the six minutes following the initial trigger. Between explicit reports, the aircraft location is interpolated for each minute and assigned a value of zero. All values are reported in 0.02-width bins.

2.3 STRATIFICATIONS

Performance results are stratified spatially, temporally, and according to certain turbulence intensity thresholds.

ALTITUDE BINS

Results are aggregated into the following altitude ranges:

Stratification	
Near-surface	0 – 9999 ft
Low	10000 – 19999 ft
Middle	20000 – 29999 ft
High	30000 – 50000 ft

Note that PIREPs and DAL EDR data are available for all altitude bins; UAL EDR data are usable only above 20000 ft.

TEMPORAL STRATIFICATION

Forecast performance in winter months (Jan–Mar) will be compared against the performance in the summer months (Jul–Sep). Performance as a function of valid time was also examined but not found to be significant.

GEOGRAPHIC STRATIFICATION

Performance is examined across four geographic regions: West, Central, Northeast, and Southeast, defined as shown in the Fig. 2.1.

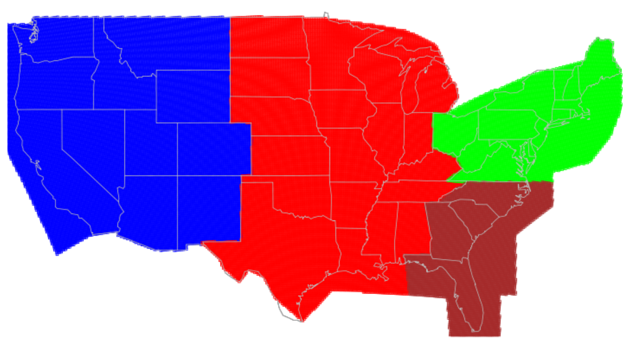


Figure 2.1: Map of the geographic regions.

INTENSITY STRATIFICATIONS

Forecast performance is also examined across a range of intensity thresholds ranging from Light to Severe turbulence (EDR values: 0.1 – 0.4+, PIREP values: 1 – 5+). The thresholds for Light, Moderate, and Severe turbulence are as follows:

Aircraft class	Light	Moderate	Severe
Light	0.13	0.16	0.36
Medium	0.15	0.20	0.44
Heavy	0.17	0.24	0.54

For all the results presented here, the Medium weight aircraft class thresholds are used.

3 APPROACH

GTGN is intended as a nowcast product to be used for tactical decisions. GTG3 provides forecasts out to 18h to support strategic decision-making. In other words, these two products are meant to provide complementary information. There is, currently, no gridded product providing real-time situational awareness for atmospheric turbulence; the best available product is a short-term GTG3 forecast. Therefore, they will be compared here as competing products.

The evaluation consists of four primary assessment areas:

1. Overall comparison between GTGN and GTG3
2. Sensitivity of performance by season: summer vs. winter
3. Sensitivity of performance by altitude layer
4. Sensitivity of performance by geographic region

Given that turbulence conditions between observations are unknown, verification of turbulence forecasts must be observation-based. That is, verification is based on the set of observations, and the forecasts are then matched to these observations. In this report, the forecasts are paired with the observations using a nearest-neighbor approach.

4 METHODS

A variety of verification approaches are employed in this assessment. They are described in the following subsections.

4.1 GTG Distributions

The characteristics of the GTGN and GTG3 fields are first evaluated using value-based distributions. Distributions are generated for each product ranging from 0 to 1.0, using a bin size of 0.01. The ratio of the GTGN and GTG3 distributions is used to facilitate comparison between the two.

4.2 Climatology/Difference Maps

Further evaluations of the characteristics of GTGN and GTG3 are performed using climatology maps. Spatial distributions are derived by aggregating counts of GTG turbulence values exceeding a threshold over a date range, issue times, and vertical layers as defined in section 2.3, for each horizontal grid point. Difference maps are generated by calculating, at each horizontal grid point, the difference in the counts, where positive values indicate a higher GTGN count and negative values indicate a higher GTG3 count.

4.3 Forecast-Observation Pairing Techniques

To enable forecast comparisons and evaluation of quality, forecasts and observations are matched spatially and temporally using the following mechanics:

- Spatially, all observations are matched to the nearest grid point in x, y, and z.
- GTGN uses observations over the previous hour, thus only observations during the 15 minutes after the GTGN issuance time are used.

GTG3 is verified using the same 15-minute accumulation of observations.

4.4 Defining Yes/No Events

For both forecasts and each of the observation types, an event is simply when the forecast or observed value exceeds a chosen threshold. A variety of thresholds are used, spanning the range from Light to Severe turbulence. Similarly, a non-event is assigned when the forecast or observed value does not exceed the chosen threshold. The one exception to this rule occurs for Light turbulence as observed by DAL EDR-equipped aircraft. Because of the triggering approach used by DAL (see section 2.2.2), the interpolated null values (i.e., the "filled in" reports at one-minute intervals that appear in the absence of a triggered or heartbeat report) can, in actuality, represent observed EDR values anywhere from zero to 0.18. As a result, interpolated nulls are ignored when evaluating Light turbulence.

4.5 Evaluations

Terminology and score definitions are first provided for reference in the subsequent sections:

MOG	Moderate-or-Greater Turbulence
POD (= PODy)	Probability of Detection: proportion of all observed events that are correctly forecast to occur, in this case, of detecting turbulence at or above a specific threshold

POFD (= 1 – PODn)	Probability of False Detection: proportion of all observed non-events that are mistakenly forecast to be events, in this case, detecting turbulence less than the specified threshold
PSS	Peirce Skill Score (aka True Skill Score, TSS): POD – POFD; Skill relative to an unbiased random forecast; provides a measure of the product's ability to separate 'yes' events from 'no'
ROC	Receiver Operating Characteristic (ROC): curve made up of (POFD, POD) pairs as the forecast threshold is varied
AUC	Area Under the (ROC) Curve: measure of ability of forecast to correctly distinguish between events and non-events
% Volume:	The percent of possible volume (the forecast domain) that is covered by the forecast for a range of threshold values

4.5.1 GTG Evaluation

Due to the non-systematic nature of the verification data set (PIREPs, and even EDR, since planes will avoid known areas of turbulence when possible), the "yes" observations and "no" observations must be treated separately (Carriere et al. 1997). As a result, it becomes inappropriate to compute several common statistics that would otherwise be computed and analyzed (e.g. Critical Success Index, Bias, and False Alarm Ratio). The rationale for this is well documented by Brown and Young (2000) and Carriere et al. (1997).

The association of the GTG product to observations as described in section 4.2 yields the following contingency table:

Hit:	forecast = yes; obs = yes
False alarm:	forecast = yes; obs = no
Miss:	forecast = no; obs = yes
Correct no:	forecast = no; obs = no

where 'yes' signifies that the forecast or observation equals or exceeds a given threshold, and 'no' signifies that the forecast or observed value is less than the threshold.

POD, POFD, and PSS are computed from the contingency table. Varying the forecast threshold for a given observation threshold produces a set of POD and POFD pairs, which form a ROC curve. Similarly, varying the forecast threshold yields a set of POD and % Volume pairs, forming a curve looking much like the ROC curve. Unlike with POFD in the ROC curve, the % Volume is independent of the observations and so a function of the forecast only.

5 Results

The presentation of the results of this evaluation will begin with overall performance followed by an examination of the performance as a function of altitude and geographic region.

5.1 Overall Performance

Before looking at the verification scores, it is useful to examine characteristics of the fields themselves, specifically distributions of the forecast values. Figure 5.1 shows distributions of turbulence intensities from both GTGN (blue) and GTG3 (red), along with the ratio of the two distributions (green). The distributions are similar, but GTGN has stronger tails: there is more very light turbulence and more strong turbulence in GTGN. The focus of the difference between the two distributions on the tails is very likely a result of the greater weight given in the GTGN algorithm to Null and MOG turbulence reports. Because turbulence observations are sparse, the number of grid points affected by these observations is tiny compared to the total number of grid points in the GTG domain. Therefore, near the peak of the distribution, the ratio of the distributions is nearly unity. By contrast, in the tails, where by definition the number of grid points is small, the effect of the observations on the ratio of the distributions is much greater.

The effect of this difference in distributions can be seen in a plot of the POD versus % volume (Fig. 5.2). More very light turbulence values for GTGN results in smaller volumes and lower POD for low forecast thresholds (upper-right portion of curves). Conversely, the greater preponderance of stronger turbulence values in GTGN translates into larger forecast volumes and higher PODs for the high forecast thresholds (lower-left portion of the curves).

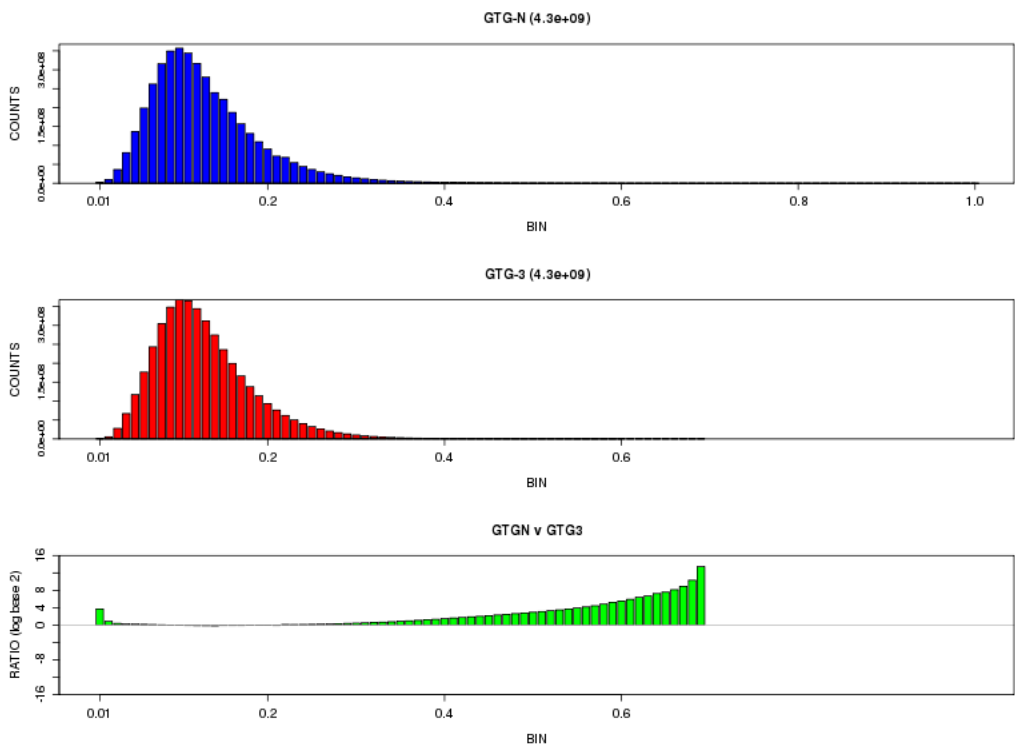


Figure 5.1: Distribution of GTGN (blue), GTG3 (red), and the ratio of GTGN to GTG3 (green) for winter. The ratio is expressed in $\log_2$, i.e., -2 denotes a ratio of 1/4.

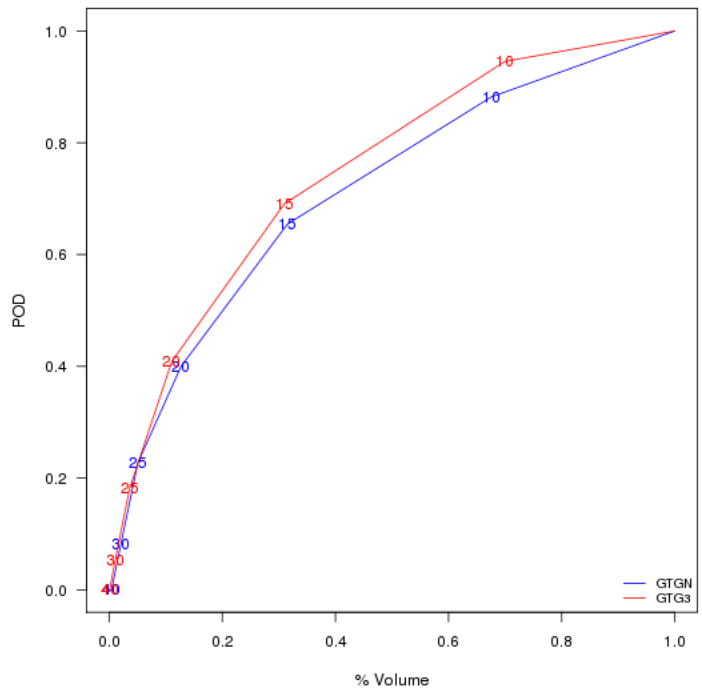


Figure 5.2: POD as a function of the % volume of the forecast for GTGN (blue) and GTG3 (red) verified against UAL EDR observations using an event threshold of 0.2 For Winter. Numbers along the curves mark various forecast thresholds (number equals threshold * 100) at their associated POD and volume.

In summer, as the jet stream shifts north and convectively-induced turbulence becomes more common than shear- and mountain wave-based turbulence, both GTGN and GTG3 have less turbulence overall, with narrower distributions that are shifted toward weaker turbulence (Fig. 5.3), when compared to winter. The differences between the GTGN and GTG3 distributions remain the same, however: more very light turbulence and more MOG turbulence in GTGN relative to GTG3.

The less numerous and weaker distributions in summer result in smaller volumes for each of the highlighted thresholds (Fig. 5.4). The smaller volumes (and lower PODs), in turn, translate into points rotated counter clockwise along the curves, pushing the crossover point above which GTGN produces larger volumes to a lower threshold (0.10 in summer compared to 0.15-0.20 in winter). Whereas in winter most of the GTGN curve lies beneath the GTG3 curve, in summer, the two curves are coincident but with the thresholds located differently along the curve. To summarize, in summer, GTG3 tends to produce smaller volumes and lower PODs compared to in winter; GTGN restores the POD to near the winter values, but does so while retaining most of the reduction in volume.

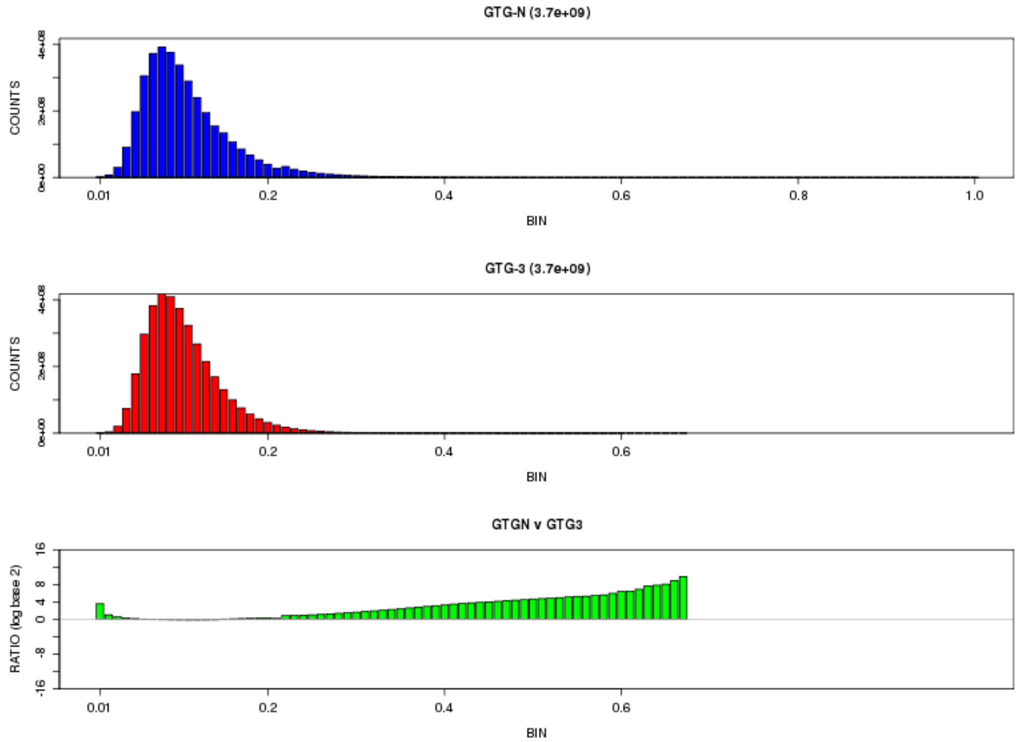


Figure 5.3: As in Fig. 5.1 but for summer.

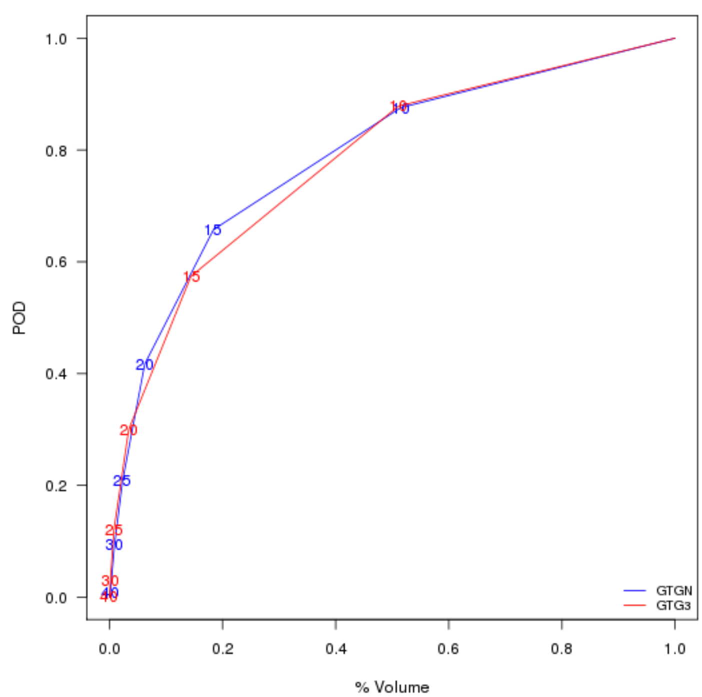


Figure 5.4: As in Fig. 5.2, but for summer.

The % Volume, plotted in the above curves, considers all forecast grid points. Most of those grid points are not changed by the GTGN algorithm. By contrast, the ROC curves draw only from forecast grid points associated with an observation. Because of the persistence of conditions likely to spawn turbulent eddies, these forecast grid points are an order-of-magnitude more likely to be adjusted by the GTGN algorithm. As a result, the expectation is that greater differences should exist in the ROC curves. When verified against DAL and UAL EDR observations and PIREPs, in winter, the GTGN and GTG3 curves are very close to each other, especially for DAL EDR (Fig. 5.5). However, even in those curves, the location of low forecast threshold points does vary considerably. In particular, GTGN produces many fewer false alarms (i.e., a lower POFD) than does GTG3 when using the 0.1 threshold (compare red and blue circles in Fig. 5.5). For higher thresholds (e.g., the red and blue triangles), the reduction in POFD is much smaller, but because of the already low false-alarm rate at these thresholds in GTG3, the percent reduction is still substantial, at least for DAL and UAL. These trends are seen more clearly in Fig. 5.6. For a forecast and observed threshold of 0.1, GTGN and GTG3 have a similar POD, but GTGN lowers the POFD considerably. As the threshold is increased, GTGN becomes increasingly superior to GTG3 in terms of POD, while the difference in false alarms disappears for the 0.3 and 0.4 thresholds. The sum of the described changes is such that the increase in skill for GTGN over GTG3 is consistent across thresholds.

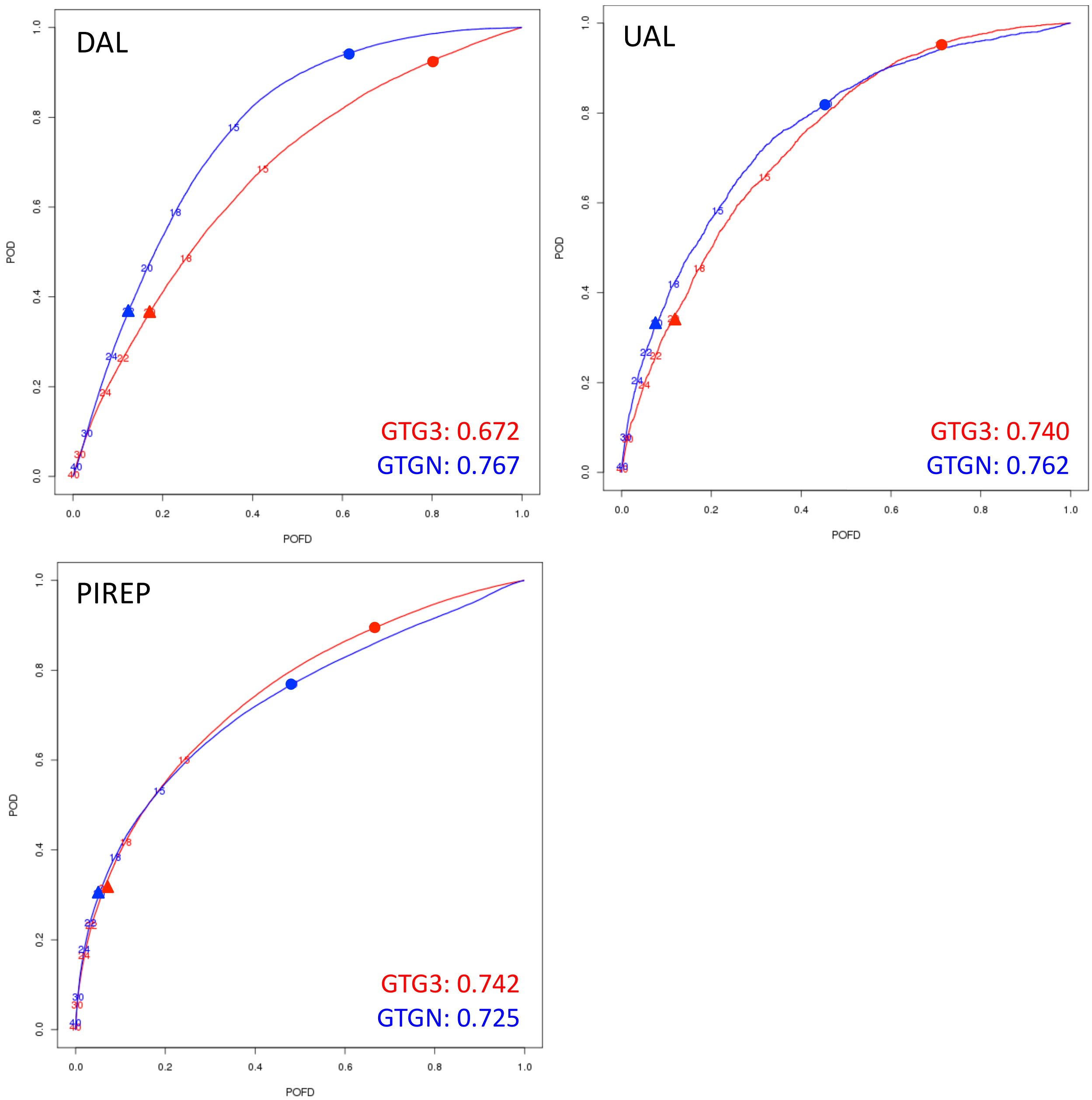


Figure 5.5: Receiver operating characteristic (ROC) curves for GTG3 (red) and GTGN (blue) for the Light (0.15) UAL (upper Right) and DAL (upper left) EDR thresholds, and Light PIREPs (lower left) For the winter Season. Area under the ROC curve (AUC) is shown in the bottom right corner. Numbers along the curves mark various forecast thresholds (number equals threshold * 100) at their associated POD and POFD. Markers highlight the location of the 0.1 (circle) and 0.2 (triangle) forecast thresholds.

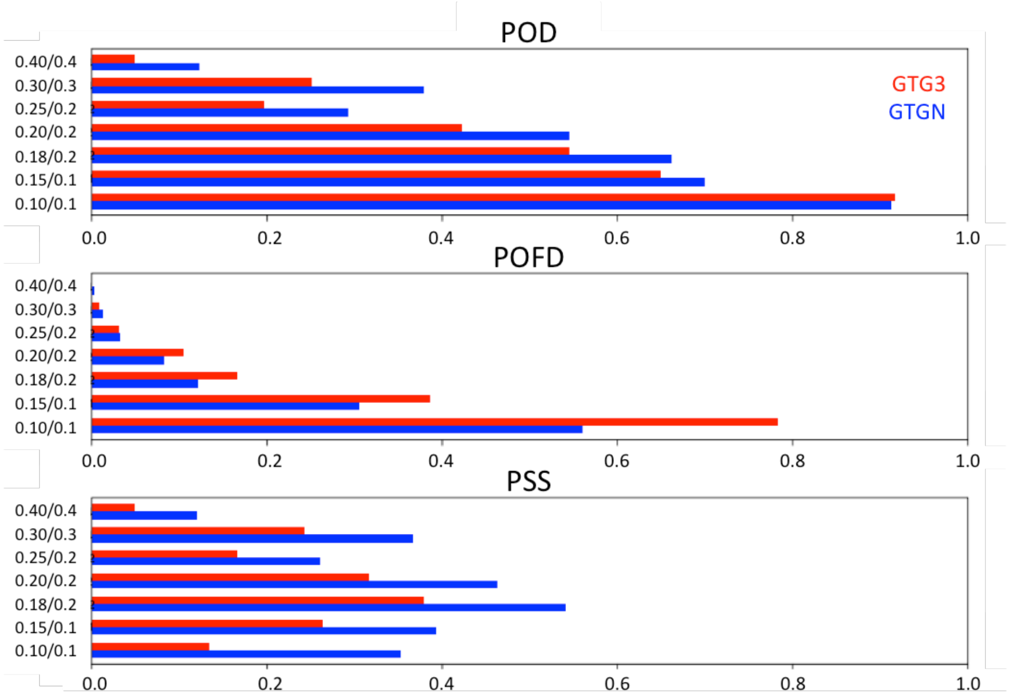


Figure 5.6: POD (top), POFD (middle), and PSS (Bottom) for GTGN (blue) and GTG3 (red) for the winter season for a set of forecast/observed thresholds as indicated by labels on the vertical axis.

In summer, there is a greater separation between the GTGN and GTG3 curves (Fig. 5.7, right). For low forecast thresholds, using GTGN would lead to fewer missed events (i.e., a higher POD) while reducing false alarms substantially. For forecast thresholds of 0.15 and above, the improvement in GTGN over GTG3 comes mainly from the reduction in missed events. With GTG3 there is a substantial reduction in forecast performance in summer (Fig. 5.7, right) compared with winter (Fig.5.7, left). GTGN, on the other hand, is able to retain most of the skill, but note that the location of the forecast thresholds rotates around the ROC curves toward the lower left. This means that in order to achieve the same level of forecast performance, it will be necessary to use a lower forecast threshold (e.g., to achieve the same performance in summer as is found using the 0.2 forecast threshold in winter, one would want to use a forecast threshold around 0.17). From the climatology maps of GTGN and GTG3 in summer (Fig. 5.8), the increased performance of GTGN over GTG3 appears to come from the additional turbulence forecast over the Great Plains and Eastern U.S., presumably a result of the NTDA algorithm.

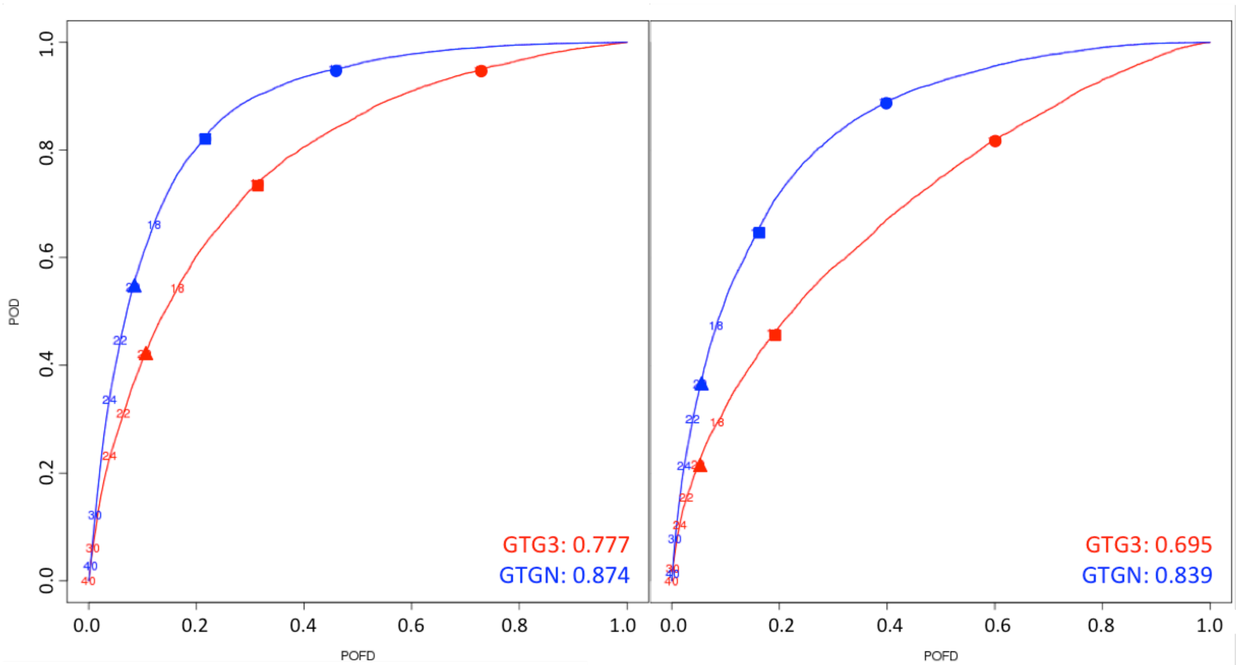


Figure 5.7: As in Fig. 5.5, but for moderate turbulence (0.2) using DAL EDR, for winter (left) and summer (right). Markers highlight the location of the 0.1 (circle), 0.15 (square), and 0.2 (triangle) forecast thresholds.

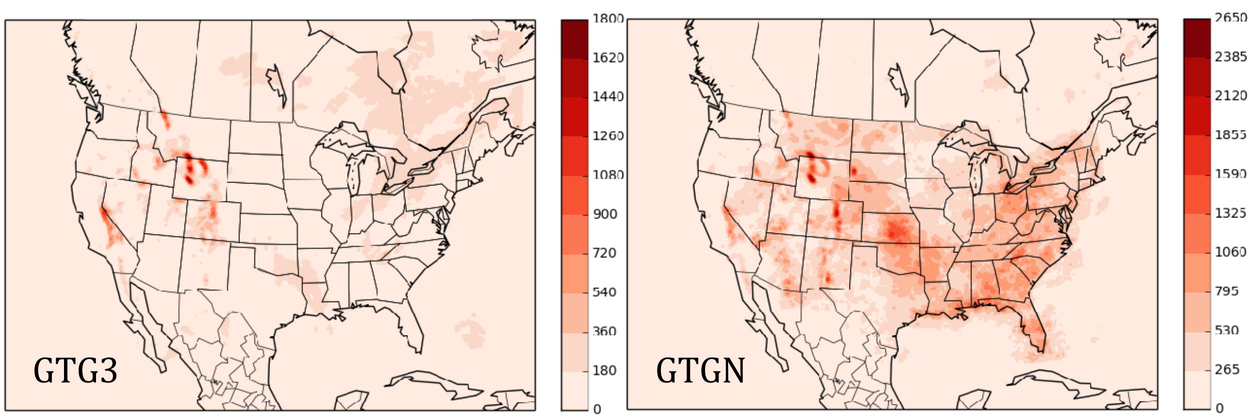


Figure 5.8: The number of times for each pixel column that a level contained a forecast of 0.25 or greater, aggregated over all forecast times over the summer period for GTG3 (left) and GTGN (Right). Scale is shown in the color bars beside each panel.

SUMMARY

GTGN has a fatter distribution than GTG3; that is, GTGN has more very low forecast values and more high values. The effect of this is that, in the winter, using GTGN in place of a 2-h GTG3 forecast would reduce the number of unforecasted turbulence encounters (i.e., missed events) while also reducing the number of false alarms. In summer, using GTGN would still reduce the number of misses—likely due to the radar-based NDTA algorithm—but with little change to the false alarms. These changes in forecast accuracy combine such that GTGN achieves better skill and superior ROC curves (i.e., a better ability to distinguish between events and non-events).

5.2 ALTITUDE AND REGIONAL SENSITIVITY

5.2.1 ALTITUDE SENSITIVITY

5.2.1.1 WINTER

Certain patterns emerge when examining the two turbulence algorithms in specific layers of the atmosphere (see section 2.3 for a definition of the layers used in this evaluation). As might be expected from the results in the previous section, when looking at a forecast threshold of 0.1 in winter, there is less GTGN at all levels (Fig. 5.9). Note the higher counts around several major airports (e.g., ATL, EWR, LAX). An investigation by altitude shows the differences between GTGN and GTG3 are more diffuse at higher altitudes and are confined to areas near major terminals at lower altitudes (not shown). One possible explanation is that given the general lack of convectively- induced turbulence in the winter, the GTGN algorithm must draw nearly exclusively on *in situ* reports. At cruising altitudes, these reports are more wide spread; at very low altitudes, these reports are necessarily focused around major airports. The greater count of GTG3 indicates that light turbulence forecast values are primarily being shifted downward by the GTGN algorithm.

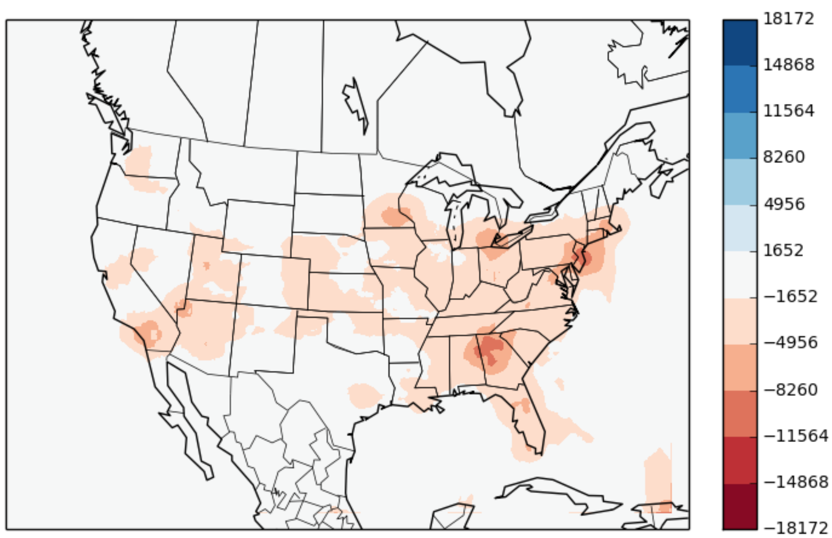


Figure 5.9: Comparison of the counts of GTGN and GTG3 greater than 0.1 over the full domain in winter. Red shading denotes aeras where GTG3 forecasts are more numerous, blue shading where GTGN forecasts are more numerous. The magnitude of the difference in counts is displayed in the color bar.

At higher forecast thresholds, a different pattern emerges. At low levels, GTGN is more numerous, especially along the East Coast (Fig. 5.10, left). Also, the differences are more widespread, not confined to areas near major airports as they were for the light turbulence forecasts. At high altitudes, GTG3 is much more numerous over the Rockies, with only small differences seen elsewhere (Fig. 5.10, right). The Turbulence PDT has noted that GTG3 tends to overpredict mountain-wave turbulence (Sharman 2014, personal communication), and so it appears that, at higher levels, at least, GTGN successfully reduces the incidence of stronger turbulence forecasts over the mountains.

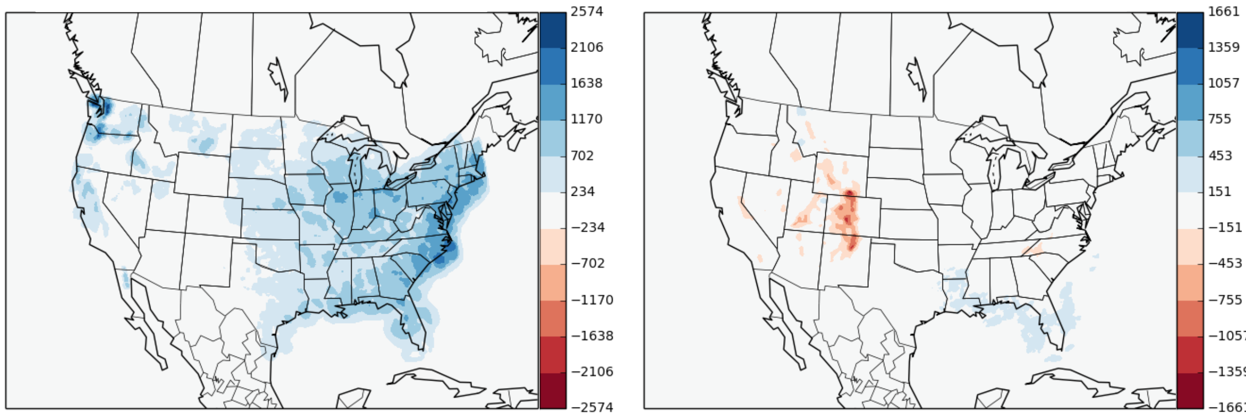


Figure 5.10: As in fig. 5.9, but for turbulence forecasts greater than 0.25 for the Near-surface layer (left) and the High layer (right) during the winter season.

The skill and accuracy of the algorithms display a similar sensitivity to the altitude layer (Fig. 5.11). For the Near-surface layer, GTGN produces a substantially higher POD (i.e., reduces missed events) at all thresholds while increasing false alarms at MOG thresholds. These combine such that GTGN is clearly more skillful at all thresholds (i.e. GTGN has higher PSS).

By contrast, in the High layer, GTGN reduces false alarms at all thresholds, but at the cost of lowering POD (i.e., increasing missed events) at thresholds ≤ 0.25. This results in GTGN being more skillful for the lighter and stronger turbulence but having similar skill to GTG3 for moderate thresholds.

The accuracy and skill plots show better performance at lower levels, or rather, greater improvement over GTG3—the performance for both algorithms is substantially better in the High layer than it is in the Near-surface layer (Fig. 5.11). Examination of the areas under the ROC curves (Fig. 5.12) confirms this conclusion; the Near-surface layer captures both the smallest area and the greatest improvement of GTGN over GTG3. Examination of the individual ROC curves for the Near-surface layer (Fig. 5.13) suggests that the improvement of GTGN over GTG3 comes solely for the Moderate or weaker forecast thresholds. Note, however, the difference in thresholds where the two curves intersect: 0.24 for GTGN, 0.20 for GTG3. So, whereas for a given POFD to the left of the intersection there does exist a GTG3 threshold that will yield a better POD than the GTGN threshold at that POFD, using the matching threshold achieves superior skill for GTGN—one with a higher POFD but a much higher POD.

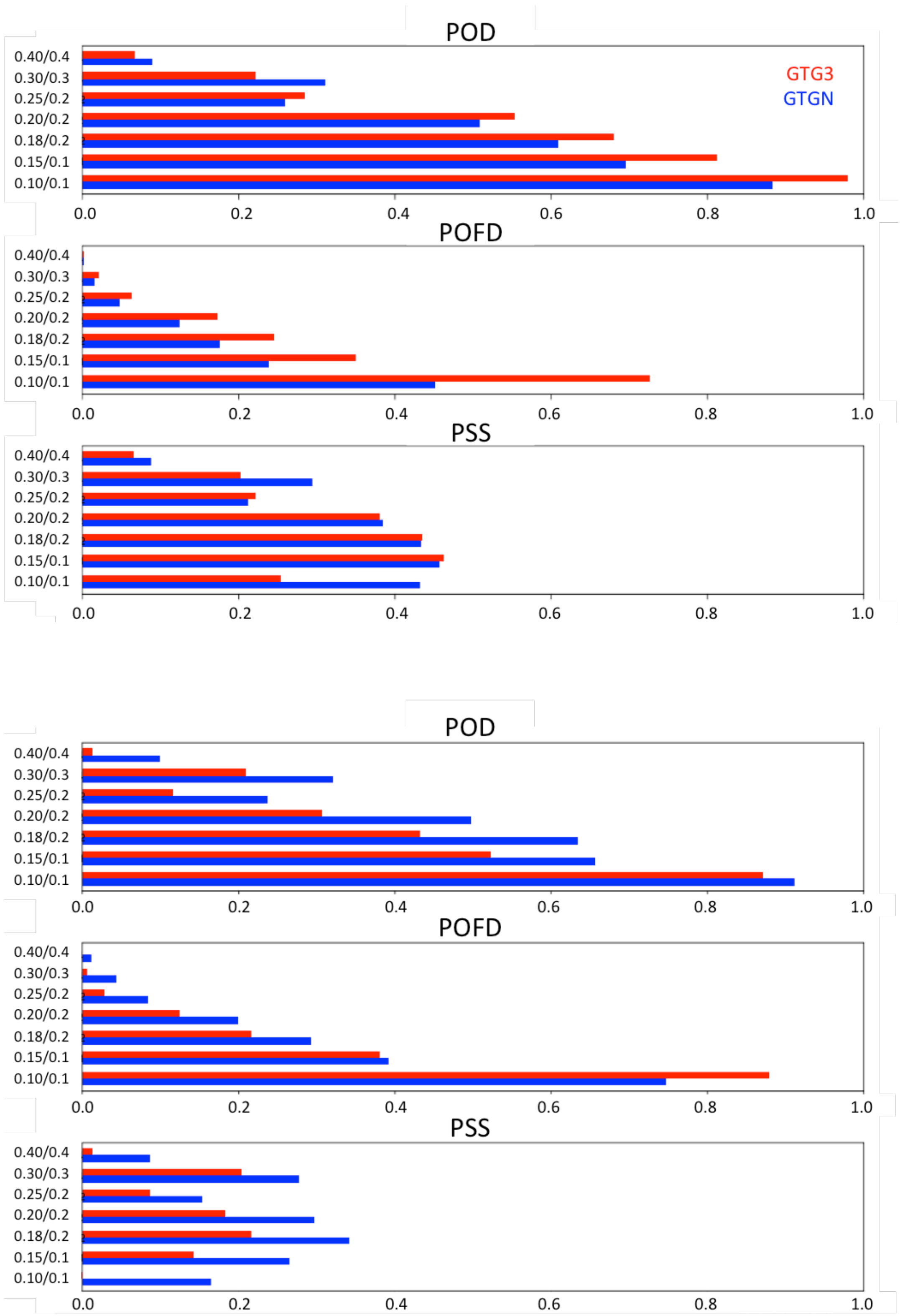


Figure 5.11: As in fig. 5.6, but for the High layer (top) and the Near-surface layer (bottom) during the Winter season.

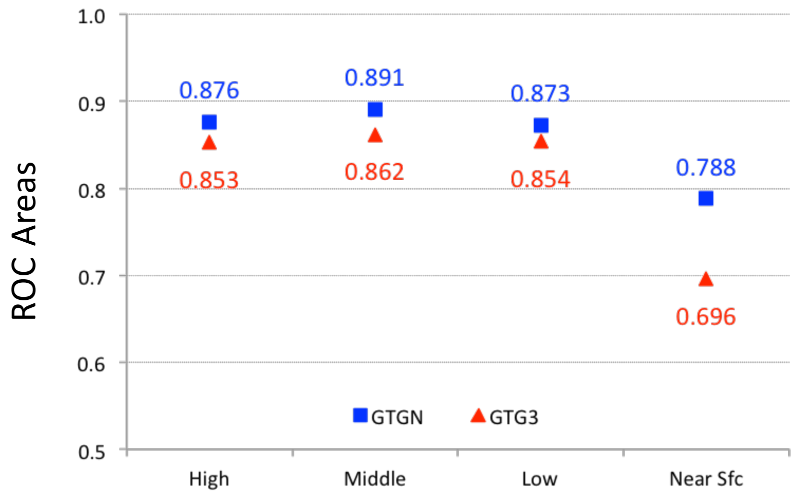


Figure 5.12: Area Under the ROC curve (AUC) for Moderate turbulence (0.2) from DAL EDR measurements for GTGN (blue squares) and GTG3 (red triangles) by altitude layer during the winter season.

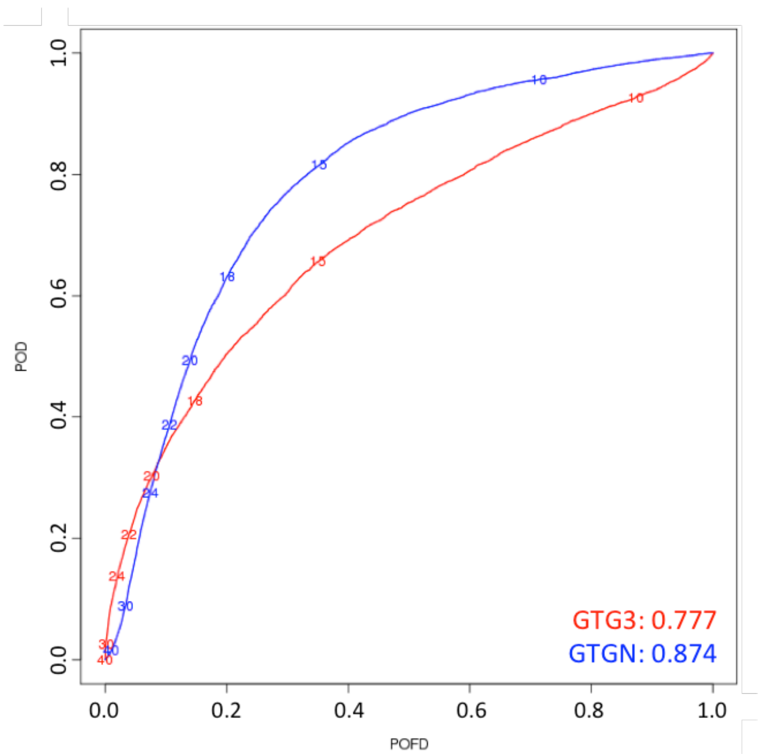


Figure 5.13: ROC curves for GTGN (blue) and GTG3 (red) for Moderate turbulence (0.2) from DAL EDR for the Near-surface layer during the winter season.

5.2.1.2 Summer

Recall that in winter, the differences between GTGN and GTG3 are confined to areas around major airports (Fig. 5.9). In summer, a very different picture emerges (Fig. 5.14, top left)): for the northern half of the U.S., along with the West Coast, GTGN is less numerous than GTG3 above the 0.1 forecast threshold, while along the Gulf Coast and southern Atlantic states, GTGN is much more numerous. However, this dipole pattern results from a vertical split. Below 30 kft, GTGN is more numerous over much of the country, with a focus on the Southeast (Fig. 5.14, bottom left), while for the cruising altitudes, GTGN is less numerous everywhere, with a focus along a corridor bounded roughly by interstates 70 and 80 (Fig. 5.14, bottom right). That is, observations tend to reduce the turbulence forecast in the shear-dominated regions near the jet stream, and increase the turbulence forecast in the thunderstorm-dominated region in the Southeast (cf. Fig. 5.15).

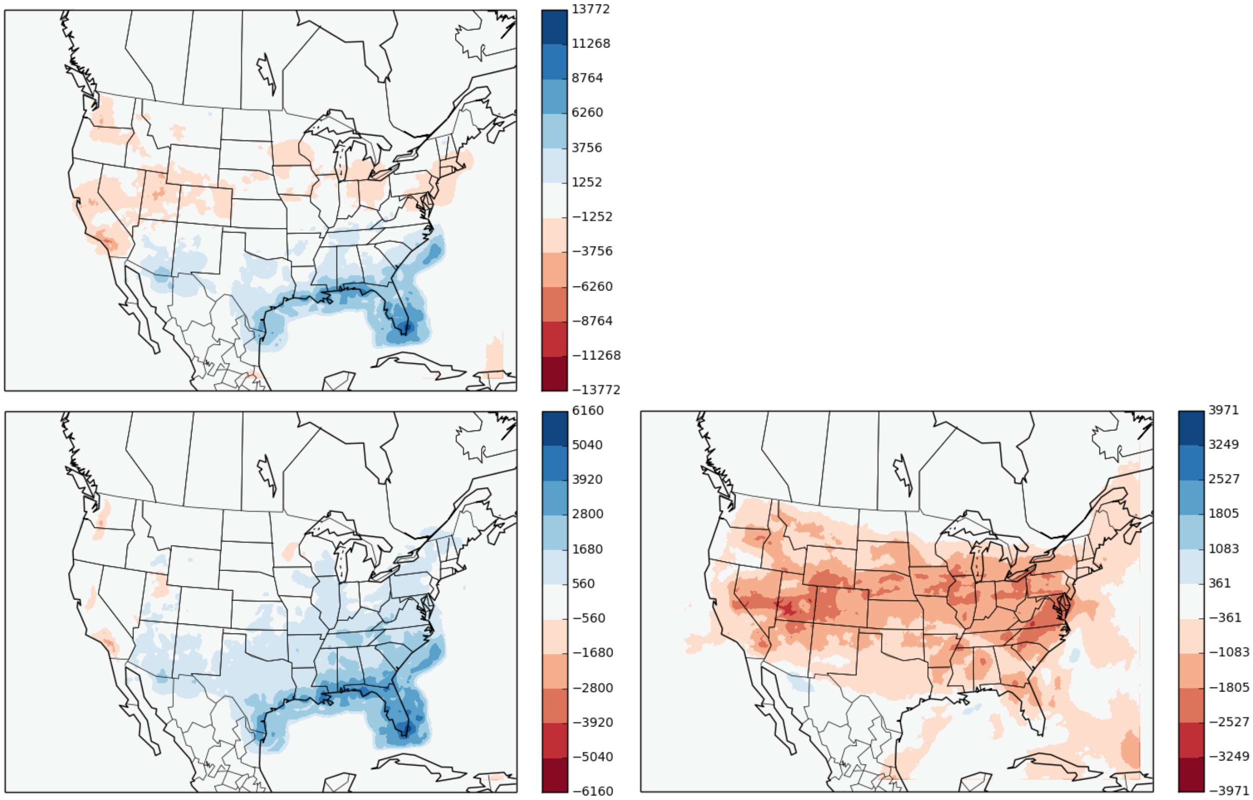


Figure 5.14: As in Fig. 5.9, but for turbulence forecasts greater than or equal to 0.1 for all layers (top left), the Low layer (bottom left), and the High layer (bottom right) during the summer season.

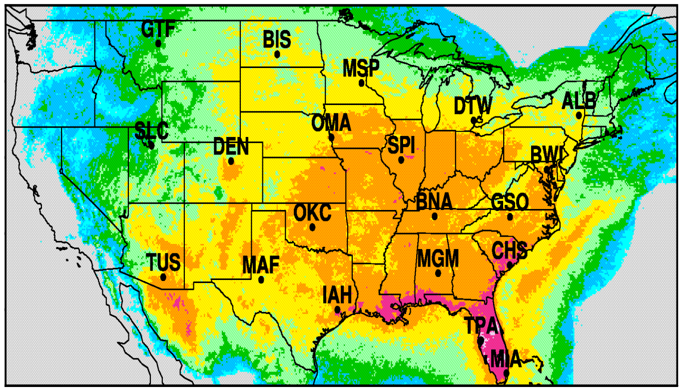
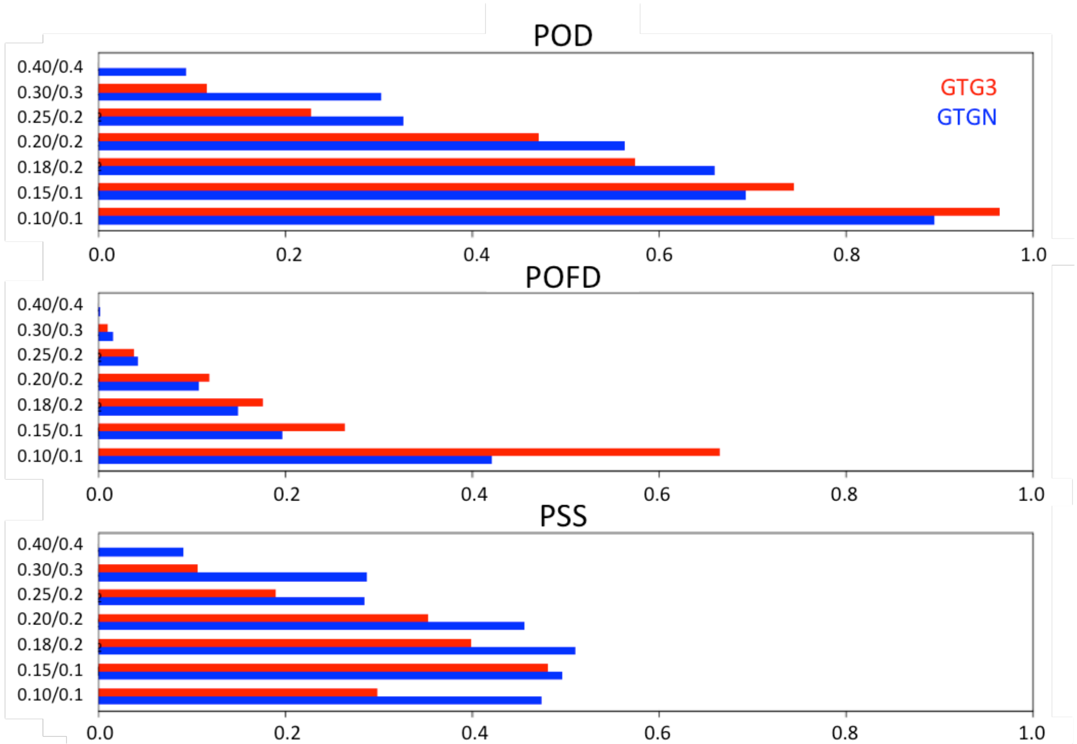


Figure 5.15: National Lightning Detection Network flash density for June, July, and August 1995-2009 (from University of Albany).

For the Winter season, GTGN improves over GTG3 by reducing false alarms at the cost of an increase in missed events (Fig. 5.11). For the Summer season (Fig. 5.16), the improvement of GTGN over GTG3 in the High layer comes from both decreasing the number of missed events for MOG (i.e., higher POD) and reducing the number of false alarms resulting in GTGN being more skillful for all thresholds. By contrast, for the Middle layer, GTGN suffers from a slight increase in false alarms, but gains a large decrease in missed events, resulting in GTGN skill scores roughly double those for GTG3. For all layers, GTGN increases skill more so in Summer than in Winter (compare Fig. 5.17 with Fig. 5.12), but the biggest increase is for the Middle layer, where the GTGN algorithms reduce the first-guess GTG3 values, indicating an overforecast bias in GTG3.



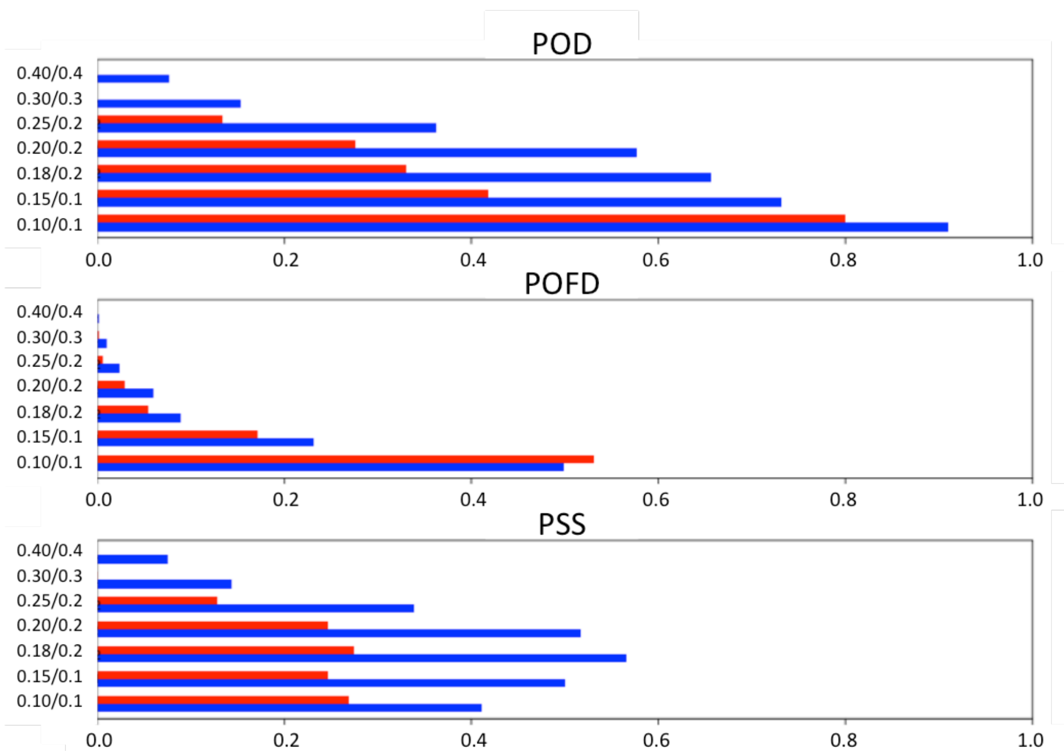


Figure 5.16: As in fig. 5.6, but for the High layer (top) and the Middle layer (bottom) during the summer season.

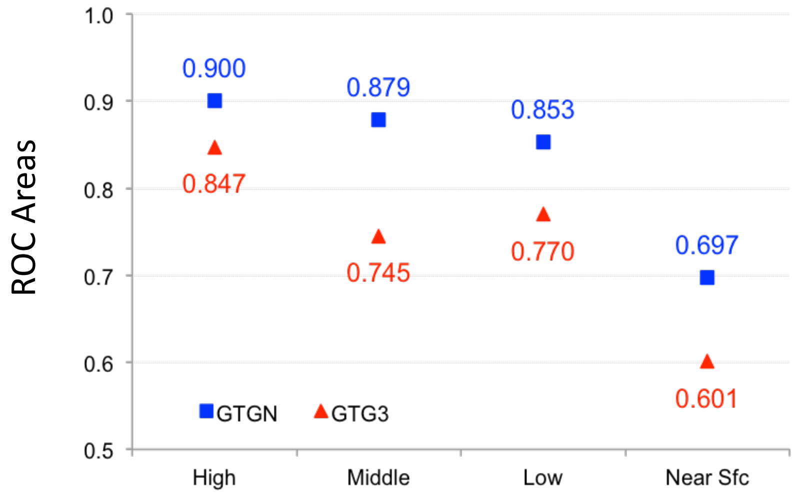


Figure 5.17: As in Fig. 5.12, but during the summer season.

5.2.2 Regional Sensitivity

The previous sections have provided some information about the differences between GTGN and GTG3, geographically, for example: the increase in MOG turbulence in GTGN in the summer, east of the Rockies (Fig. 5.8); for forecast thresholds of 0.1 or greater, GTGN is less numerous in the winter over the eastern half of the country (Fig. 5.9); in the summer, GTGN is much more numerous in the Southeast (Fig. 5.14). Taken altogether, GTGN gives superior performance over all regions, but especially over the Southeast (Fig. 5.18). The improvement in performance is greatest in the summer, particularly in the Central and Northeast regions. This highlights the significant role of the NTDA component of the GTGN algorithm.

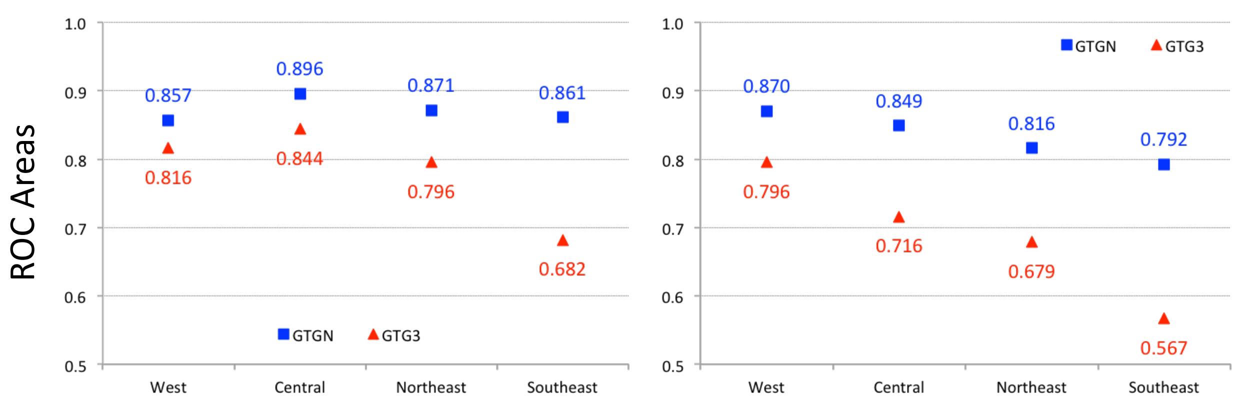


Figure 5.18: As in Fig. 5.12, but by region for winter (left) and summer (right).

SUMMARY

In winter, GTGN turbulence is more common than in GTG3 for lower flight levels, but less common than in GTG3 for higher flight levels. This reduces the number of missed events in low levels, but at the cost of more false alarms. In contrast, for higher levels, GTGN reduces the number of false alarms (relative to GTG3), but captures more turbulent events. GTGN outperforms GTG3 for all levels, but the improvement is greatest in the 0 – 10 kft layer. GTGN tends to reduce the number of strong turbulence events over the Rockies, while increasing the number of low-level turbulence events over the eastern half of the country, such that the improvement in forecast performance is greatest over the Southeast.

In summer, Light turbulence is less common in GTGN than in GTG3, except across the Southeast, but MOG turbulence is more common throughout all layers and across all regions. In lower layers, this reduces the number of missed events at the cost of more false alarms, but for higher flight levels, both false alarms and missed events are reduced. GTGN improvement over GTG3 is greater in the summer compared to winter and largest over the Southeast, where thunderstorms are most numerous.

5.3 PERSISTENCE

GTGN forecasts are provided every 15 minutes, drawing upon observations over the preceding hour, with greater weight given to more recent observations. Running the GTGN algorithm is computationally expensive, and it may provide useful information beyond the 15-minute period for which it is designed (If GTGN can make use of observations up to an hour old, it is reasonable to think that the nowcast could still provide value up to an hour in the future.) To address this question, a brief examination of the performance of GTGN persistence is provided. As expected, performance does decline as one moves from the 0–15 minute valid period through to the 45–60 minute valid period, however, the drop in performance, as measured by the area under the ROC curve, is much less than the improvement in the GTGN forecast over GTG3 (Fig. 5. 19).

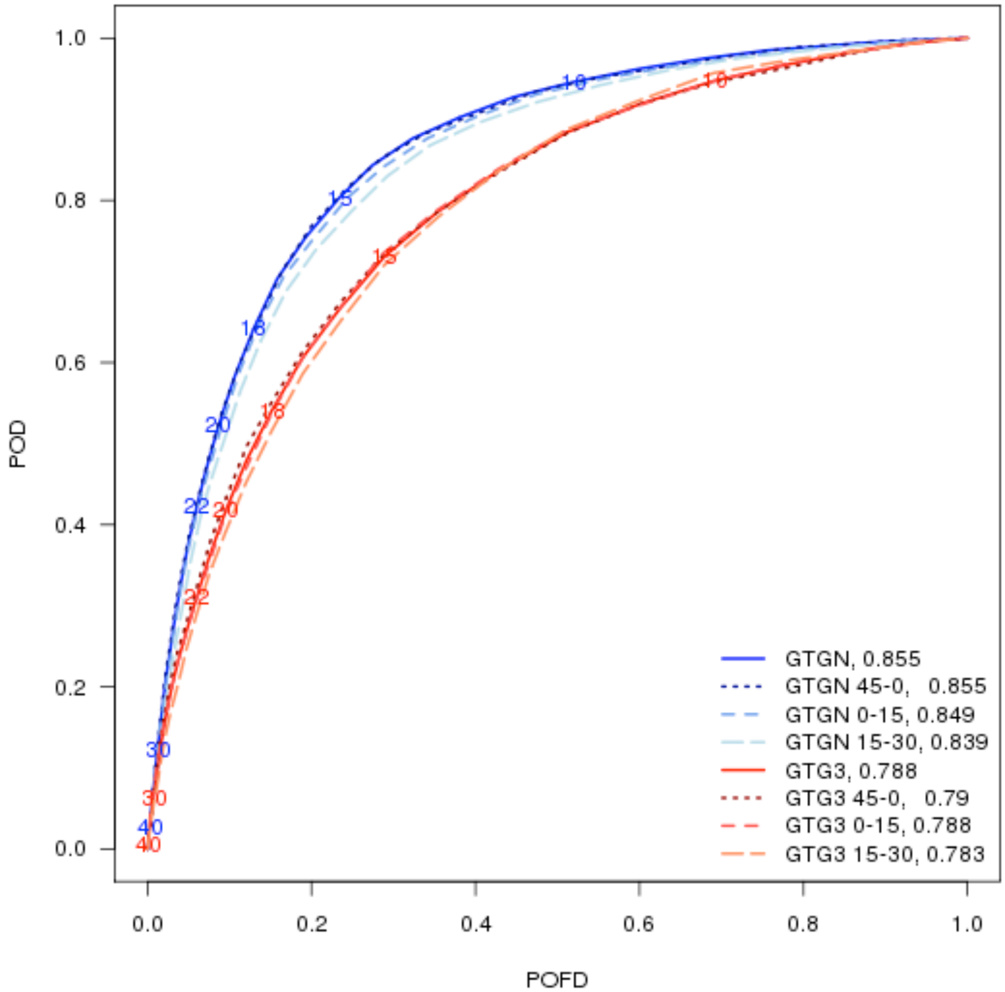


Figure 5.19: ROC curves for GTGN (blue) and GTG3 (red) for leads of 0–15 minutes (solid), 15–30 minutes (dotted), 30–45 minutes (short dash), and 45–60 minutes (long dash).

6 SUMMARY

The GTGN product is designed to provide near-real-time situational awareness to support operational aviation turbulence decisions. GTGN uses a short-term (1- or 2-h lead) forecast from the GTG3 product as a first-guess field which is then augmented with direct observations—both pilot reports (PIREPs) and EDR measurements—and remotely sensed data from the NEXRAD Turbulence Detection Algorithm (NTDA). Updates are supplied every 15 minutes, with output provided on the same Rapid Refresh (RAP)-based, 13-km grid used by GTG3.

GTGN has a fatter distribution than GTG3; that is, GTGN has more very low forecast values and more high values. The effect of this is that, in the winter, using GTGN in place of a 2-h GTG3 forecast would reduce the number of unforecasted turbulence encounters (i.e., missed events) while also reducing the number of false alarms. In summer, using GTGN would still reduce the number of misses—likely due to the radar-based NDTA algorithm—but with little change to the false alarms. These changes in forecast accuracy combine such that GTGN achieves better skill and superior ROC curves (i.e., a better ability to distinguish between events and non-events).

In winter, GTGN turbulence is more common than in GTG3 for lower flight levels, but less common than in GTG3 for higher flight levels. This results in a reduction in missed events in low levels, but at the cost of more false alarms. In contrast, for higher levels, GTGN reduces the number of false alarms (relative to GTG3), but captures more turbulent events. GTGN outperforms GTG3 for all levels, but the improvement is greatest in the 0 – 10 kft layer. GTGN tends to reduce the number of strong turbulence events over the Rockies, while increasing the number of low-level turbulence events over the eastern half of the country, such that the improvement in forecast performance is greatest over the Southeast.

In summer, Light turbulence is less common in GTGN than in GTG3, except across the Southeast, but MOG turbulence is more common throughout all layers and across all regions. In lower layers, the result is a reduction in missed events at the cost of more false alarms, but for higher flight levels, both false alarms and missed events are reduced. GTGN improvement over GTG3 is greater in the summer compared to winter and largest over the Southeast, where thunderstorms are most numerous.

GTGN is intended as a nowcast product to be used for tactical decisions. GTG3 provides forecasts out to 18h to support strategic decision-making. In other words, these two products are meant to provide complementary information. There is, currently, no gridded product providing real-time situational awareness for atmospheric turbulence; the best available product is a short-term GTG3 forecast. Therefore, they are compared here as competing products. In this context, the inclusion of recent observations in the GTGN product yields consistently better forecast performance when compared to GTG3.

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. The authors would like to thank the Turbulence Product Development Team for providing the forecast data that was needed for the evaluation.

References

Brown, B.G., and G.S. Young, 2000: Verification of icing and icing forecasts: Why some verification statistics can't be computed using PIREPs. Preprints, 9th conference on Aviation, Range, and Aerospace Meteorology, Orlando, FL, Sep. 11-15, American Meteorological Society (Boston), 393-398.

Carriere, J.M., S. Alquier, C. LeBot, and E. Moulin, 1997: Statistical verification of forecast icing risk indices, *Meteor. Appl.,* **4,** 115-130.

NWS, 2007: Aviation Weather Services, Advisory Circular AC 00-45F. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, and U.S. Department of Transportation, 393 pp.

University of Albany, 2016: Verification of Daily Thunderstorm Occurrence Using the National Lightning Detection Network (NLDN). http://www.atmos.albany.edu/facstaff/kristen/tsclimo/tsclimo.html