

Evaluations of Upper-Level Turbulence Diagnostics Performance Using the Graphical Turbulence Guidance (GTG) System and Pilot Reports (PIREPs) over East Asia

JUNG-HOON KIM AND HYE-YEONG CHUN

Department of Atmospheric Sciences, Yonsei University, Seoul, South Korea

ROBERT D. SHARMAN AND TEDDIE L. KELLER

Research Application Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 25 November 2010, in final form 26 February 2011)

ABSTRACT

The forecast skill of upper-level turbulence diagnostics is evaluated using available turbulence observations [viz., pilot reports (PIREPs)] over East Asia. The six years (2003–08) of PIREPs used in this study include null, light, and moderate-or-greater intensity categories. The turbulence diagnostics used are a subset of indices in the Graphical Turbulence Guidance (GTG) system. To investigate the optimal performance of the component GTG diagnostics and GTG combinations over East Asia, various statistical evaluations and sensitivity tests are performed. To examine the dependency of the GTG system on the operational numerical weather prediction (NWP) model, the GTG system is applied to both the Regional Data Assimilation and Prediction System (RDAPS) analysis data and Global Forecasting System (GFS) analysis and forecast data with 30-km and 0.3125° (T382) horizontal grid spacings. The dependency of the temporal variation in the PIREP and GFS data and the forecast lead time of the GFS-based GTG combination are also investigated. It is found that the forecasting performance of the GTG system varies with year and season according to the annual and seasonal variations in the large-scale atmospheric conditions over the East Asia region. The wintertime GTG skill is the highest, because most GTG component diagnostics are related to jet streams and upper-level fronts. The GTG skill improves as the number of PIREP samples and the vertical resolution of the underlying NWP analysis data increase, and the GTG performance decreases as the forecast lead time increases from 0 to 12 h.

1. Introduction

Upper-level turbulence continues to be a hazard for the commercial aviation industry, especially at cruising altitudes at which passengers and crew are more likely to be unbuckled (Lester 1994). A significant portion of upper-level turbulence encounters is of the clear-air turbulence variety (Wolff and Sharman 2008; Kim and Chun 2011); these are especially difficult to avoid because they occur unexpectedly without visual indicators such as on-board radar echoes and visible clouds. According to the

2009 annual report of the National Transportation Safety Board (NTSB 2009), turbulence was the leading cause of weather-related aircraft accidents from 1996 to 2005. It was recently suggested that turbulence may have been a potential contributing factor for the Air France flight 447 disaster over the Atlantic Ocean on 1 June 2009 (Kaplan and Vollmer 2010). As air transportation density has increased, forecasting of upper-level turbulence has become more important for both aviation safety and reducing economical cost.

Turbulence forecasting skill has improved with developing science and engineering technologies. Today, the most promising methods for turbulence forecasting or avoidance are as follows. First, the pilot may try to avoid turbulent regions identified by pilot reports (PIREPs) from previous aircraft encounters. Second, aircraft can attempt to strategically avoid turbulence by using empirical turbulence forecasting techniques or advisories [e.g., Airmen's Meteorological Information (AIRMET),

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Prof. Hye-Yeong Chun, Dept. of Atmospheric Sciences, Yonsei University, 262 Seongsanno, Seodaemun-gu, Seoul 120-749, South Korea.
E-mail: chunhy@yonsei.ac.kr

Significant Meteorological Information (SIGMET), and Federal Aviation Administration (FAA) guidelines for thunderstorm avoidance (Lane et al. 2003)]. An example of an empirical forecasting technique is satellite-based observations of transverse cloud bands (Knox et al. 2010; Lenz et al. 2009) or lee-wave signatures (Uhlenbrock et al. 2007; Feltz et al. 2009), which may be used to infer turbulence potential. Third, as computer capacity has increased, locally focused high-resolution numerical modeling has been used to explicitly predict aircraft-scale turbulence, especially over mountain regions (e.g., Clark et al. 2000; Olafsson and Agustsson 2009; Lane et al. 2009; Kim and Chun 2010). Fourth, numerical weather prediction (NWP) model output can be used as input to automated turbulence diagnostics to infer upper-level turbulence potential (Sharman et al. 2006).

From the meteorological perspective, turbulence forecasting by explicitly resolving the small-scale turbulent eddies that affect commercial aircraft (roughly 10–1000 m) is still not possible given the current grid spacings (~10–20 km) of operational NWP models. Turbulence potential may be estimated from current NWP systems on the basis of the assumption that the energy associated with aircraft-scale turbulent eddies cascades down from the resolved large-scale atmospheric disturbances, however (e.g., Dutton and Panofsky 1970; Koshyk and Hamilton 2001; Cho and Lindborg 2001; Tung and Orlando 2003). This is the basis of the fourth approach listed above. There are many possible NWP-based turbulence diagnostics or indicators, however, and an integrated approach in which several such indicators are used in combination as a consensus seems to provide better overall performance than the use of just one diagnostic (Sharman et al. 2006). This integrated approach to upper-level turbulence prediction is termed the Graphical Turbulence Guidance (GTG; Sharman et al. 2006), and output from the GTG system is currently available operationally (online at <http://aviationweather.gov/adds>).

The approach used for upper-level turbulence forecasting in the GTG system is as follows. First, the GTG system calculates several turbulence diagnostics representing large-scale forcings such as frontogenesis, tropopause proximity, and ageostrophic flow. Second, the calculated diagnostics are optimally combined by weighting scores that are based on the forecasting performance of the individual diagnostics. To achieve economically efficient and operationally useful forecasting performance of the GTG system [the recommended probabilities of moderate-or-greater (MOG) and null (NIL) turbulence detection are greater than 0.8 and 0.85, respectively (Sharman et al. 2006)], newly developed turbulence diagnostics {e.g., eddy dissipation rate [Eq. (3.7) in Frehlich and Sharman (2004)], Lighthill–Ford radiation

[Eq. (23) in Knox et al. (2008)]} and various statistical combination methods (e.g., Abernethy 2008) have been continually evaluated and implemented in updated versions of the GTG system. A global GTG system that is based on the Global Forecast System (GFS) forecasting data has recently been developed and is currently being evaluated (Williams et al. 2010).

According to climatologically calculated turbulence diagnostics obtained from the various operational models [e.g., the Aviation global model (AVN; Ellrod et al. 2003) and the 40-yr European Centre for Medium-Range Weather Forecasts Re-Analysis data (ERA-40; Jaeger and Sprenger 2007)], the highest potential for upper-level turbulence is located over the East Asia region where jet streams are strongest (Koch et al. 2006). Here, the East Asia region is the domain including South Korea, eastern China, and Japan (as shown in Fig. 2b, described below). The current study focuses on the evaluation of a GTG system over East Asia using two different operational NWP model systems [the Regional Data Assimilation and Prediction System (RDAPS) and GFS] and available PIREPs. Sensitivity tests that are based on annually and seasonally distributed PIREPs are conducted to determine the optimal performance of the GTG system over East Asia. These investigations can provide useful information for the evaluation of the global GTG system over East Asia and could provide valuable information for pilots, dispatchers, and forecasters to reduce upper-level turbulence encounters and maintain air-flight safety over East Asia. This work also extends that of Sharman et al. (2006) by exploring the performance sensitivity to the number of turbulence diagnostics used in the GTG combination, seasonal dependencies, and dependencies due to the underlying NWP model, temporal variation in the PIREP and NWP model data, and forecast lead time.

The remainder of this paper is organized as follows. In section 2, the data and method of the evaluations are described. The GTG system used in this study is described in section 3. In section 4, the results of evaluations and several sensitivity tests of the GTG system that use six years of PIREP data over East Asia are provided. Comparison between the RDAPS-based and GFS-based GTG systems and sensitivity tests to the temporal variations in the PIREP and GFS data and forecast lead time of the GTG system are also presented in section 4. Summary and conclusions are provided in section 5.

2. Data and method

a. PIREPs

At this time, verbal PIREPs used in this study are the only routinely available observations of upper-level

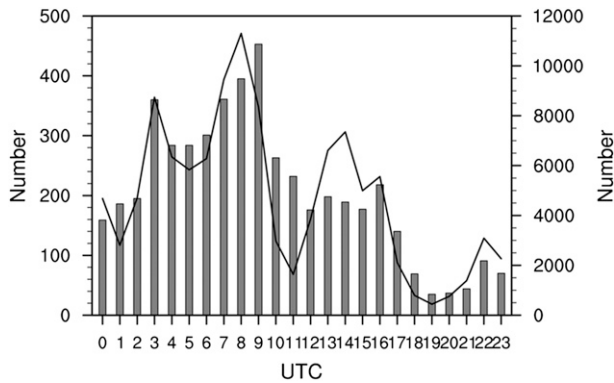


FIG. 1. Hourly distribution of the number of total (solid line; right axis) and MOG-level (bars; left axis) PIREPs at upper levels (above 20 000 ft) over East Asia for six years (2003–08).

turbulence over East Asia. The PIREP data have been accumulated at both the National Center for Atmospheric Research (NCAR) and the Korea Aviation Meteorological Agency for six years from December 2002 to November 2008 (hereinafter 2003–08). Because the detailed collection processes of PIREP data and their inherent subjectivities have already been examined in previous studies (e.g., Sharman et al. 2006; Kim and Chun 2011), we only briefly describe the PIREP data used in this study. The PIREPs include information about upper-level turbulence encounters such as intensity, altitude, location, and time. Turbulence intensity as reported has nine classifications [0–9, where 0, 1–2, 3–4, 5–6, 7, and 9 correspond to NIL, light (LGT), moderate (MOD), severe (SEV), extreme (EXT), and missing, respectively]. To isolate upper-level turbulence, turbulence encounters that occurred only above 20 000 ft (FL200; ~6100 m) including some multilayer reports (e.g., FL200–FL240) are considered. Figure 1 shows the hourly distribution of the number of total and MOG-level PIREPs at upper levels (above FL200) over East Asia for the 6-yr period. In Fig. 1, large diurnal variations in the number of observations are evident, with daytime (0000–1100 UTC) numbers larger than those in the nighttime (1200–2300 UTC).

To evaluate the GTG system, the PIREP data collected over the East Asia region within ± 2 h of 0000 and 1200 UTC are used. The total number of PIREPs used is 30 911, and includes NIL (18 701), LGT (9970), and MOG (1370) turbulence reports. Because the reported LGT-level events are very uncertain and the number of SEV- and EXT-level events are simply too small to construct reliable statistics, we only consider the NIL-versus-MOG discrimination capability of the GTG system (e.g., Sharman et al. 2006; Kim et al. 2009). Figure 2 shows the monthly and spatial distributions of the MOG-level turbulence encounters for the six years (2003–08). The

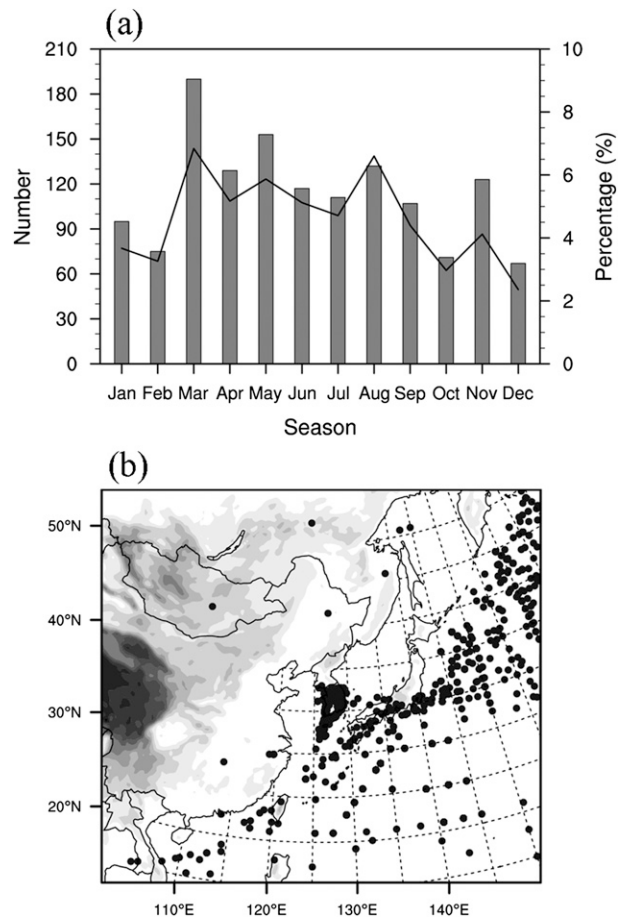


FIG. 2. (a) Monthly and (b) spatial distributions of the MOG-level turbulence encounters over East Asia, occurring within ± 2 h of 0000 and 1200 UTC for six years (2003–08). The solid line in (a) depicts relative percentages of the MOG-level turbulence normalized by the total turbulence within a given month.

MOG-level events occurred more frequently in the spring and summer seasons than in the autumn and winter seasons (Fig. 2a). In the spatial distribution, the MOG-level events are dominant along the flight routes between the southern and eastern Asia regions and the North American continent (Fig. 2b).

b. Evaluation method

The evaluation method used in this study is based on the computation of the probability of detection (POD; e.g., Mason 1982). In this study, the NIL- and MOG-level events are exclusively used to avoid the intensity uncertainty of the LGT-level events (e.g., Sharman et al. 2006; Kim et al. 2009). Two PODs are constructed: probability of detection of “yes” (PODY) for the MOG-level events and the probability of detection of “no” (PODN) for the NIL-level events as

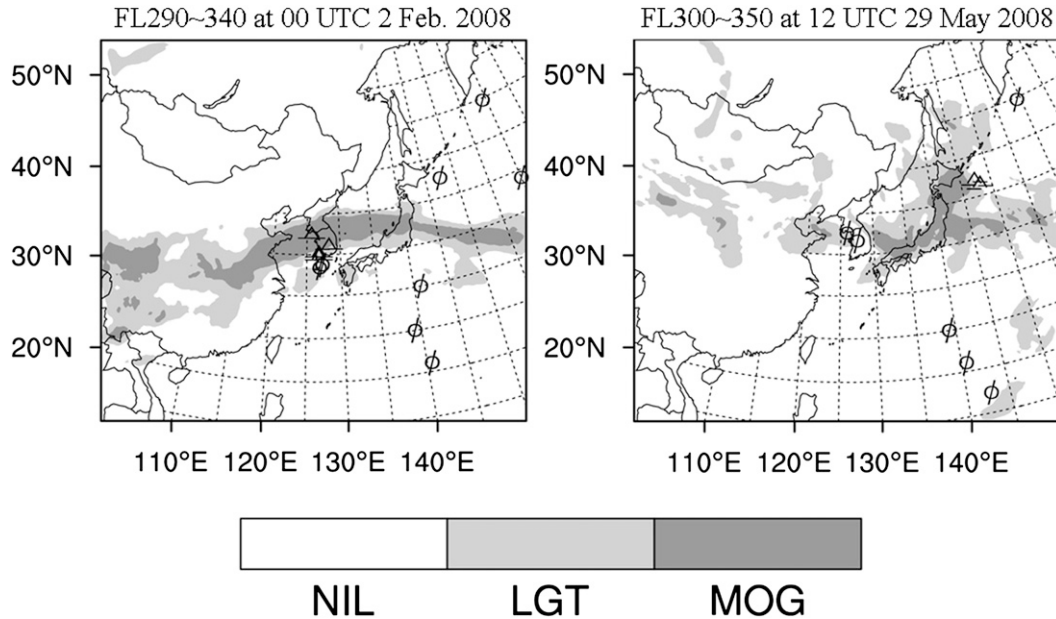


FIG. 3. Examples of the 2008 GTG_{EA} at (left) $z = 29\,000\text{--}34\,000$ ft (FL290–340) for 0000 UTC 2 Feb 2008 and (right) $z = 30\,000\text{--}34\,000$ ft (FL300–350) for 1200 UTC 29 May 2008 on the RDAPS domain, derived using the selected 20 diagnostics and thresholds listed in Table 1, on the basis of the 6-yr PIREP data. Several MOG- and NIL-level PIREPs observed within ± 2 h of the targeted times are depicted using conventional symbols for turbulence intensity. The thresholds for the boundaries between NIL and LGT levels and LGT and MOG levels are 0.3 and 0.475, respectively.

$$\begin{aligned}
 \text{PODY} &= \frac{Y_{\text{obs}} Y_{\text{for}}}{Y_{\text{obs}} Y_{\text{for}} + Y_{\text{obs}} N_{\text{for}}} \quad \text{and} \\
 \text{PODN} &= \frac{N_{\text{obs}} N_{\text{for}}}{N_{\text{obs}} Y_{\text{for}} + N_{\text{obs}} N_{\text{for}}}. \quad (1)
 \end{aligned}$$

In Eq. (1), Y and N mean “yes” and “no”, respectively, and the subscripts “obs” and “for” indicate the observed (as reported in the PIREP) turbulence intensity and the forecast value of the GTG product or individual diagnostic, respectively. Note that the GTG product or individual diagnostic is calculated at the closest grid point to the PIREP. After applying this logic to all MOG- and NIL-level events, one PODY value and one PODN value are obtained for a given threshold value. When this metric is iterated through 30 given thresholds that range from the minimum to the maximum values of the GTG product or individual diagnostics, 30 PODY and PODN statistics are obtained. When the 30 PODY and PODN statistics are depicted in an x - y plot (e.g., Fig. 4, described below), where the x and y axes are the PODN and PODY, respectively, the performance of the GTG product or individual diagnostic can be measured by the area under the curve (AUC). If the value of AUC is 1, the performance is perfect (i.e., the GTG product or individual diagnostic can perfectly discriminate all MOG- and NIL-level events; Sharman et al.

2006). This method has been used in many previous evaluations of upper-level turbulence derived using various model-based diagnostics (e.g., Brown et al. 2000; Tebaldi et al. 2002; Lee et al. 2003; Sharman et al. 2006; Knox et al. 2008; Jang et al. 2009; Kim et al. 2009; Ellrod and Knox 2010).

3. The GTG system

The GTG system described in Sharman et al. (2006) provides a flexible framework for computing turbulence potential, but the system must be tuned for optimal performance because of regional dependencies of the turbulence sources and varying performance characteristics of the underlying NWP model. Thus it is necessary to reevaluate the diagnostics and relative weights used within the GTG combination for optimal performance over East Asia. Here, the GTG system using a set of climatologically weighted diagnostics (GTGC) is applied to both the RDAPS analysis data and GFS analysis and forecast data with 30-km and 0.3125° (T382) horizontal grid spacings. The model data are produced daily at 0000 and 1200 UTC, and the model domains are focused on the East Asia region as shown in Fig. 3.

The detailed procedure for constructing the GTGC combination consists of five overall steps. In the first step, the GTG system calculates a suite of diagnostics

TABLE 1. Thresholds (T1, T2, T3, T4, and T5) corresponding to null, light, moderate, severe, and extreme intensities of upper-level turbulence and AUC values for the GTG_{EA} combination and its 20 component diagnostics that are based on the 6-yr (2003–08) PIREP data over East Asia. The column labeled “units” refers to the mathematical units for the individual diagnostics; PVU is potential vorticity unit.

Index	Units	T1	T2	T3	T4	T5	AUC
GTG _{EA}	—	0	0.3	0.475	0.75	1	0.795
DTF3	m ² s ⁻²	0.4	0.6	1.1	1.8	2.5	0.781
CP	kt ²	9.5	20	52	100	150	0.780
-Ri _e	—	-20	-2.0	-0.6	-0.3	0.5	0.767
T11	s ⁻²	1.9 × 10 ⁻⁷	2.4 × 10 ⁻⁷	4.6 × 10 ⁻⁷	7.2 × 10 ⁻⁷	9.9 × 10 ⁻⁷	0.756
Brown2	cm ² s ⁻³	0.17	0.24	0.63	1.1	1.5	0.754
F _θ	m ² s ⁻³ K ⁻²	2.4 × 10 ⁻⁵	4.4 × 10 ⁻⁵	9.0 × 10 ⁻⁵	1.7 × 10 ⁻⁴	2.5 × 10 ⁻⁴	0.743
TROPV	m ⁻¹	1.5 × 10 ⁻⁴	4.5 × 10 ⁻⁴	7.4 × 10 ⁻⁴	1.0 × 10 ⁻³	1.6 × 10 ⁻³	0.742
Dutton	m s ⁻¹ km ⁻¹	20	23	32	45	58	0.711
CCAT	s ⁻²	1.6 × 10 ⁻⁸	2.3 × 10 ⁻⁸	3.0 × 10 ⁻⁸	6.3 × 10 ⁻⁸	1.3 × 10 ⁻⁷	0.691
EDRavg	m ^{2/3} s ⁻¹	0.008	0.01	0.02	0.04	0.07	0.690
Endlich	° s ⁻¹	6.9 × 10 ⁻³	7.4 × 10 ⁻³	8.0 × 10 ⁻³	2.8 × 10 ⁻²	7.4 × 10 ⁻²	0.675
HTG	K m ⁻¹	6.7 × 10 ⁻⁶	8.5 × 10 ⁻⁶	1.0 × 10 ⁻⁵	1.7 × 10 ⁻⁵	2.8 × 10 ⁻⁵	0.674
NVA	m ² s ⁻²	1.5 × 10 ⁻⁹	1.7 × 10 ⁻⁹	2.8 × 10 ⁻⁹	6.7 × 10 ⁻⁹	1.1 × 10 ⁻⁸	0.671
NCSU2	s ⁻³	2.8 × 10 ⁻¹³	3.2 × 10 ⁻¹³	5.3 × 10 ⁻¹³	1.0 × 10 ⁻¹²	1.5 × 10 ⁻¹²	0.663
NGM2	K m ⁻¹ s ⁻¹	2.5 × 10 ⁻⁷	3.4 × 10 ⁻⁷	3.7 × 10 ⁻⁷	5.8 × 10 ⁻⁷	8.0 × 10 ⁻⁷	0.647
LHF	s ⁻³	4.4 × 10 ⁻¹³	4.7 × 10 ⁻¹³	7.6 × 10 ⁻¹³	1.6 × 10 ⁻¹²	2.4 × 10 ⁻¹²	0.643
IAwind	m s ⁻¹	32	35	48	85	122	0.633
HPG	PVU m ⁻¹	4.9 × 10 ⁻⁶	5.9 × 10 ⁻⁶	6.9 × 10 ⁻⁶	1.5 × 10 ⁻⁵	2.8 × 10 ⁻⁵	0.623
RoL	—	0.14	0.15	0.2	0.36	0.51	0.599
UBF	s ⁻²	4.1 × 10 ⁻⁹	4.4 × 10 ⁻⁹	4.7 × 10 ⁻⁹	8.8 × 10 ⁻⁹	1.6 × 10 ⁻⁸	0.581

over the entire 3D grid of the input NWP model at the desired forecast time. The formulations of most diagnostics used in this study [e.g., T11 (Ellrod and Knapp 1992) and Brown2 (Brown 1973)] are described in appendix A of Sharman et al. (2006) and references therein. In this study we have also included the spontaneous inertial-gravity wave imbalance formulation, represented by the Lighthill-Ford radiation term in the Knox et al. (2008), and it is labeled LHF in subsequent tables.

In the second step, to allow direct comparisons of the individual diagnostics with the PIREP data, the calculated diagnostics in the native coordinates of the operational NWP model are vertically interpolated to the conventional flight altitudes above FL200. Note that the conventional flight altitude or flight level (FL) is an isobaric surface that is based on the assumption of the standard atmosphere (Lane et al. 2003; Sharman et al. 2006; Kim and Chun 2010), whereas the native coordinate of the RDAPS analysis data is a pressure coordinate with 22 isobaric layers from the 1000- to the 100-hPa levels.

Because the units and numerical magnitudes of the calculated diagnostics are different from each other, piecewise linear functions are used to normalize the calculated values of the individual diagnostics to common-scale values from 0 to 1 before combining, where 0 and 1 correspond to the NIL- and EXT-level events, respectively (Sharman et al. 2006). This is the third step. The piecewise linear functions (e.g., Fig. 3 in Sharman et al. 2006) for the individual diagnostics are derived using thresholds that

correspond to the median values of the accumulated individual diagnostics near all NIL-, LGT-, and MOD-level events over East Asia during the given period. The SEV- and EXT-level thresholds instead used the 98th- and 99th-percentile values of the individual diagnostics, respectively, because the number of these events is too small to calculate the median values (Kim et al. 2009).

The fourth step involves selecting the set of optimal diagnostics. To do this, the performance of each diagnostic, represented by the AUC, is calculated using the observed MOG- and NIL-level PIREP data. Table 1 shows the NIL-, LGT-, MOD-, SEV-, and EXT-level thresholds (T1, T2, T3, T4, and T5, respectively) and AUC values for the 20 best diagnostics on the basis of the six years of PIREP data over East Asia. Using the data in this table, the normalized weighting score W_n for each diagnostic n is calculated as

$$W_n = \frac{AUC_n^2}{\sum_{n=1}^N AUC_n^2}, \quad n = 1, 2, 3, \dots, N. \quad (2)$$

Note that the thresholds (T1, T2, T3, T4, and T5) and performance (AUC) of the individual diagnostics can and do depend on time of year and season and the NWP model used in the GTG system. This sensitivity will be evaluated in the next section.

In the fifth and last step, the normalized value of each diagnostic D_n^* computed at each grid point (i, j, k) of the

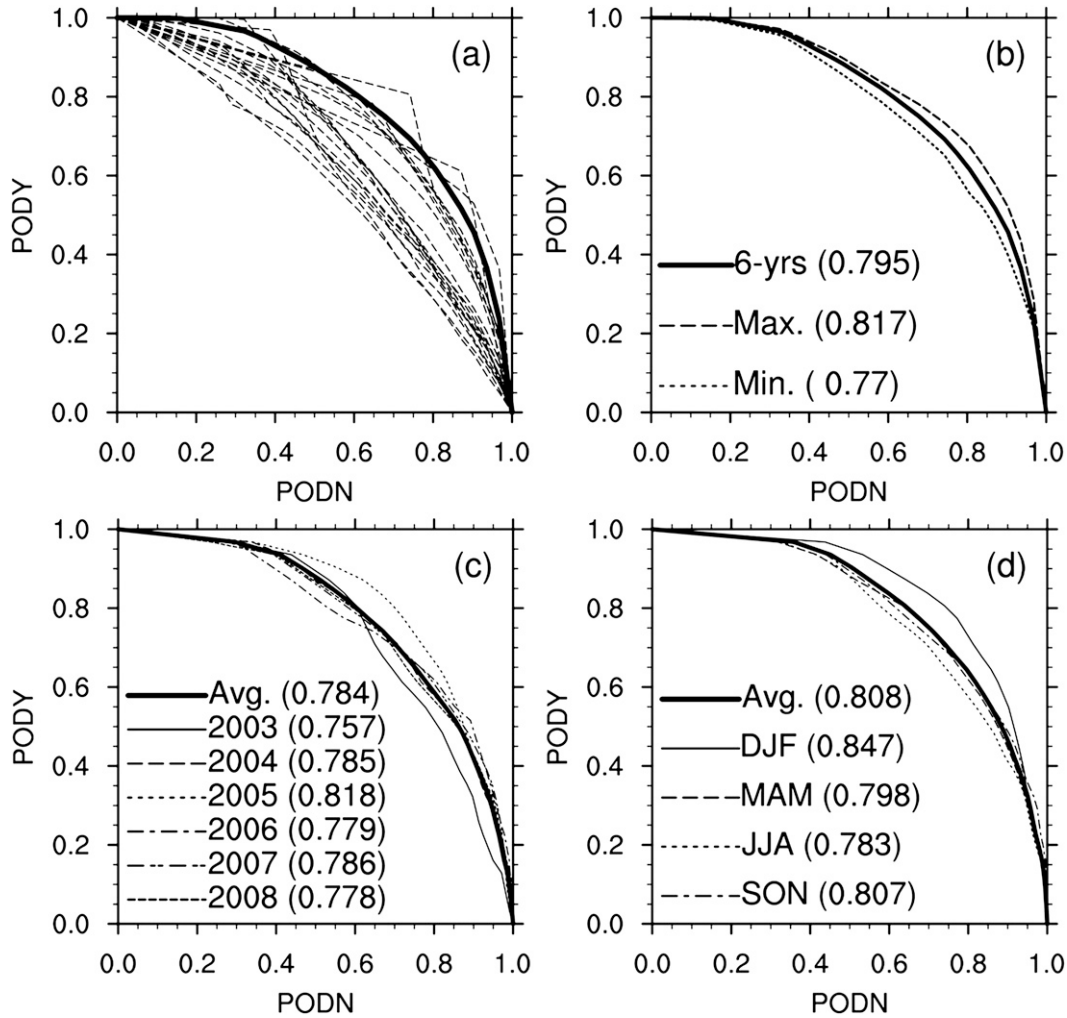


FIG. 4. PODY–PODN statistics of (a) the 6-yr GTG_{EA} (thick solid line) and 20 individual diagnostics (thin dashed lines), (b) the maximum and minimum boundaries of the 200 experiments using subsets of randomly selected half-fraction samples, and the (c) yearly and (d) seasonal GTG_{EA} , on the basis of the 6-yr PIREP and RDAPS analysis data over East Asia. In (c), plots of the 6-yr average (Avg.), 2003, 2004, 2005, 2006, 2007, and 2008 GTG_{EA} are provided. In (d), plots of the 6-yr average (Avg.), DJF, MAM, JJA, and SON GTG_{EA} are shown. The AUC values of all GTG_{EA} experiments are written in parentheses in (b)–(d).

NWP model are combined into the GTG product by using Eq. (3) with the normalized weighting scores W_n from Eq. (2). Hereinafter, the final product of the GTG in Eq. (3) is written with an EA subscript to avoid confusion with the conterminous United States (CONUS) version of the GTG system:

$$GTG_{EA}(i, j, k) = \sum_{n=1}^N W_n D_n^*(i, j, k), \quad n = 1, 2, 3, \dots, N. \tag{3}$$

Figure 3 shows examples of the 2008 GTG_{EA} at FL290–340 for 0000 UTC 2 February 2008 (left panel)

and at FL300–350 for 1200 UTC 29 May 2008 (right panel), derived using the selected 20 diagnostics and thresholds listed in Table 1. Several MOG- and NIL-level PIREPs observed within ± 2 h of the targeted times are depicted using conventional symbols for turbulence intensities. Given the large PIREP time window used for comparison, the agreement in both cases is fairly good.

4. Evaluations of the GTG_{EA} system over East Asia

In this section various statistical evaluations of the GTG_{EA} over the East Asia region are performed using both the RDAPS and GFS analysis data and the 6-yr

PIREP data. For RDAPS, results that are based on weights and thresholds determined using PIREPs from all six years combined are presented in section 4a. Using the 20 best diagnostics, the GTG_{EA} forecast performance is evaluated over the entire 6-yr period (“6-yr GTG_{EA}”), over individual years (“yearly GTG_{EA}”), and by season (“seasonal GTG_{EA}”). Comparison between the RDAPS- and GFS-based GTG_{EA} is provided in section 4b. Sensitivity tests to the temporal variations in the PIREP and GFS data (section 4c) and forecast lead time (section 4d) of the GFS-based GTG_{EA} are also investigated.

a. GTG_{EA} performance using the 6-yr RDAPS analysis and PIREP data

Figure 4a shows the PODY–PODN performance statistics of the 6-yr GTG_{EA} and its 20 component diagnostics listed in Table 1 on the basis of the 6-yr PIREP data. Note that the GTG_{EA} combination is the best performer with an AUC of 0.795. Among the individual diagnostics listed in Table 1, the overall performance of the diagnostic turbulent kinetic energy (TKE) formulation (DTF3; Marroquin 1998) index (AUC = 0.781) is the highest, and the Colson–Panofsky (CP; Colson and Panofsky 1965) index (AUC = 0.780) is the second highest. Table 2 shows other performance metrics of the 6-yr GTG_{EA} and its 20 component diagnostics listed in Table 1. In Table 2, root-mean-square error (RMSE) of the 6-yr GTG_{EA} is 0.1952, which is smaller than any of the 20 individual diagnostics. In addition, true skill score (TSS = PODY + PODN – 1 = 0.3621) of the 6-yr GTG_{EA} is higher than that of the 20 component diagnostics because both the PODY (0.6088) and PODN (0.7533) statistics of the 6-yr GTG_{EA} are higher than those of most individual diagnostics. On the basis of these results, it is concluded that the 6-yr GTG_{EA} combination provides superior performance for upper-level turbulence over East Asia from 2003 to 2008.

To address the irregular nature of PIREPs’ frequency and location, 200 additional experiments were performed as conducted in Sharman et al. (2006). From the full set of PIREP–GTG_{EA} forecast data pairs, these subsets were generated by randomly resampling only one-half of the full set of the pairs to reevaluate the 6-yr GTG_{EA}. Figure 4b shows the maximum and minimum boundaries of the 200 experiments as a function of PODY–PODN performance statistics. In Fig. 4b, results from the 200 experiments are consistent with each other within about ±3%. This stable and robust result is similar to that found in Sharman et al. (2006): ±2% for upper-level turbulence in the United States.

Figures 4c and 4d show the PODY–PODN performance statistics of the yearly and seasonal GTG_{EA} on

TABLE 2. Statistical performances for the GTG_{EA} combination and its 20 component diagnostics derived using the 6-yr (2003–08) PIREP data over East Asia and RDAPS analyses data (0000 and 1200 UTC) with 30-km horizontal grid spacing. The RMSE, PODY, PODN, and TSS scores for both the GTG_{EA} combination and individual diagnostics are calculated using the thresholds shown in Table 1. The weighting values are derived from Eq. (2) in the text.

Index	Units	RMSE	PODY	PODN	TSS	Weighting
GTG _{EA}	—	0.1952	0.6088	0.7533	0.3621	—
DTF3	m ² s ⁻²	0.2684	0.5562	0.7580	0.3142	0.063
CP	kt ²	0.2544	0.5540	0.7427	0.2967	0.063
–Ri _e	—	0.2263	0.5645	0.7336	0.2981	0.061
TI1	s ⁻²	0.2749	0.6175	0.7151	0.3326	0.059
Brown2	cm ² s ⁻³	0.2613	0.5380	0.7660	0.3040	0.059
F _θ	m ² s ⁻³ K ⁻²	0.2569	0.5208	0.7588	0.2796	0.057
TROPV	m ⁻¹	0.2516	0.4321	0.7399	0.1720	0.057
Dutton	m s ⁻¹ km ⁻¹	0.2821	0.5693	0.7050	0.2742	0.053
CCAT	s ⁻²	0.3041	0.4007	0.7470	0.1477	0.050
EDRavg	m ^{2/3} s ⁻¹	0.2950	0.4635	0.7358	0.1993	0.049
Endlich	° s ⁻¹	0.3010	0.3628	0.7640	0.1268	0.047
HTG	K m ⁻¹	0.3288	0.5037	0.6890	0.1925	0.047
NVA	m ² s ⁻²	0.3189	0.3942	0.7257	0.1198	0.047
NCSU2	s ⁻³	0.3160	0.4847	0.6862	0.1709	0.046
NGM2	K m ⁻¹ s ⁻¹	0.3561	0.4759	0.6713	0.1472	0.043
LHF	s ⁻³	0.3179	0.4416	0.6938	0.1354	0.043
IAwind	m s ⁻¹	0.3412	0.4978	0.6368	0.1346	0.042
HPG	PVU m ⁻¹	0.3376	0.3766	0.6951	0.0717	0.040
RoL	—	0.3713	0.5453	0.5726	0.1179	0.037
UBF	s ⁻²	0.3349	0.3693	0.6902	0.0595	0.035

the basis of weights determined by the 6-yr PIREP data. The GTG_{EA} performance does depend on year and season, which is likely because of the interannual and seasonal variations in the large-scale atmospheric conditions that affect the aircraft-scale turbulence over East Asia. In Fig. 4c, the 2005 GTG_{EA} (AUC = 0.818) is the best among the yearly GTG_{EA} performance values, whereas the 2003 GTG_{EA} (AUC = 0.757) is the worst. With regard to the seasonal variation (Fig. 4d), the wintertime GTG_{EA} (AUC = 0.847) is the best among the seasonal GTG_{EA} performance values, and the summertime GTG_{EA} (AUC = 0.783) is the worst. Variance in the yearly GTG_{EA} performance is 0.061, which is slightly lower than that in the seasonal GTG_{EA} (0.064). The 6-yr mean performance of the yearly GTG_{EA} is 0.784, which is lower than that of the seasonal GTG_{EA} (0.808). Among all GTG_{EA} experiments that are based on the 6-yr PIREP data, the wintertime GTG_{EA} (0.847) is the highest during this period. Higher wintertime skill is likely due to the fact that most of the component turbulence diagnostics included in the current GTG system are related to enhanced shears related to the jet stream, which are usually strong in the winter over East Asia (Kim et al. 2009). On the other hand, the summer skill is lower than for the other seasons, implying that new

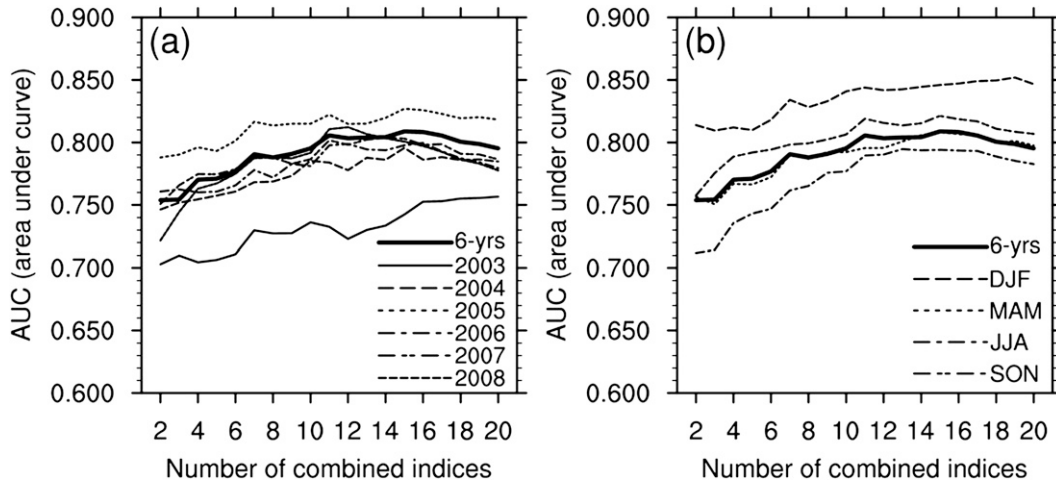


FIG. 5. AUCs of the 6-yr GTG_{EA} and (a) yearly GTG_{EA} and (b) seasonal GTG_{EA} that are based on the 6-yr PIREP and RDAPS analysis data over East Asia, as a function of the number of combined indices included in the GTG_{EA} combination. In (a), plots of the 6-yr and yearly GTG_{EA} are shown. In (b), plots of the 6-yr, DJF, MAM, JJA, and SON GTG_{EA} are plotted.

turbulence diagnostics that better represent the turbulence potential during the summertime need to be developed and implemented in the current GTG system.

When experiments for each year of the GTG_{EA} are additionally conducted using 200 subsets of randomly selected half-fraction samples for each year of the classified PIREP data, the additional 200 experiments provide performance AUCs that are consistent with each other within about ±6%–9%. For the seasonal experiments, the uncertainty boundaries are about ±4%–8%. Differences in uncertainty boundaries for each 6-yr (±3%), yearly (±6%–9%), and seasonal (±4%–8%) GTG_{EA} that is based on the 6-yr PIREP data depend on the number of selected PIREP samples. As the number of classified PIREP samples becomes larger, the uncertainty boundaries decreased, which is consistent with the results shown in Sharman et al. (2006). This shows that the interannual and seasonal dependencies of the GTG_{EA} system over East Asia are statistically significant.

Experiments were performed to assess the sensitivity of GTG_{EA} performance to the number of diagnostics included in the GTG combination. Figure 5 shows the AUC values of the 6-yr, yearly, and seasonal GTG_{EA} as a function of the number of combined indices listed in Tables 1 and 2. The performance of the 6-yr GTG_{EA} (thick solid lines) becomes higher as the number of combined indices increases up to about 15 (AUC = 0.809) and then lowers slightly, showing that the 6-yr GTG_{EA} is optimal when about 15 component diagnostics are combined. In this and similar assessments to be presented, the order of the indices added is based on the relative AUC of the index.

For the yearly GTG_{EA} (Fig. 5a), the AUC values for 2005 are always higher than the other yearly GTG_{EA} experiments, whereas those for 2003 are always lower than the others. The yearly AUC values generally become higher as the number of combined indices increases up to 12–16. The optimal AUC values for 2004, 2005, 2006, 2007, and 2008 are 0.796, 0.827, 0.800, 0.804, and 0.812 with 15, 15, 16, 14, and 12 combined indices, respectively.

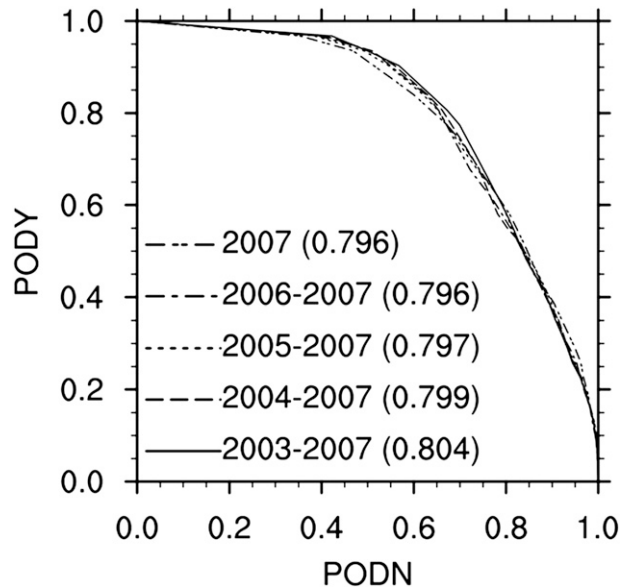


FIG. 6. PODY–PODN statistics of the 2008 GTG_{EA} that are based on the previous 1-yr (2007), 2-yr (2006–07), 3-yr (2005–07), 4-yr (2004–07), and 5-yr (2003–07) PIREP and RDAPS analysis data over East Asia. The AUC values for each GTG_{EA} combination are written in parentheses.

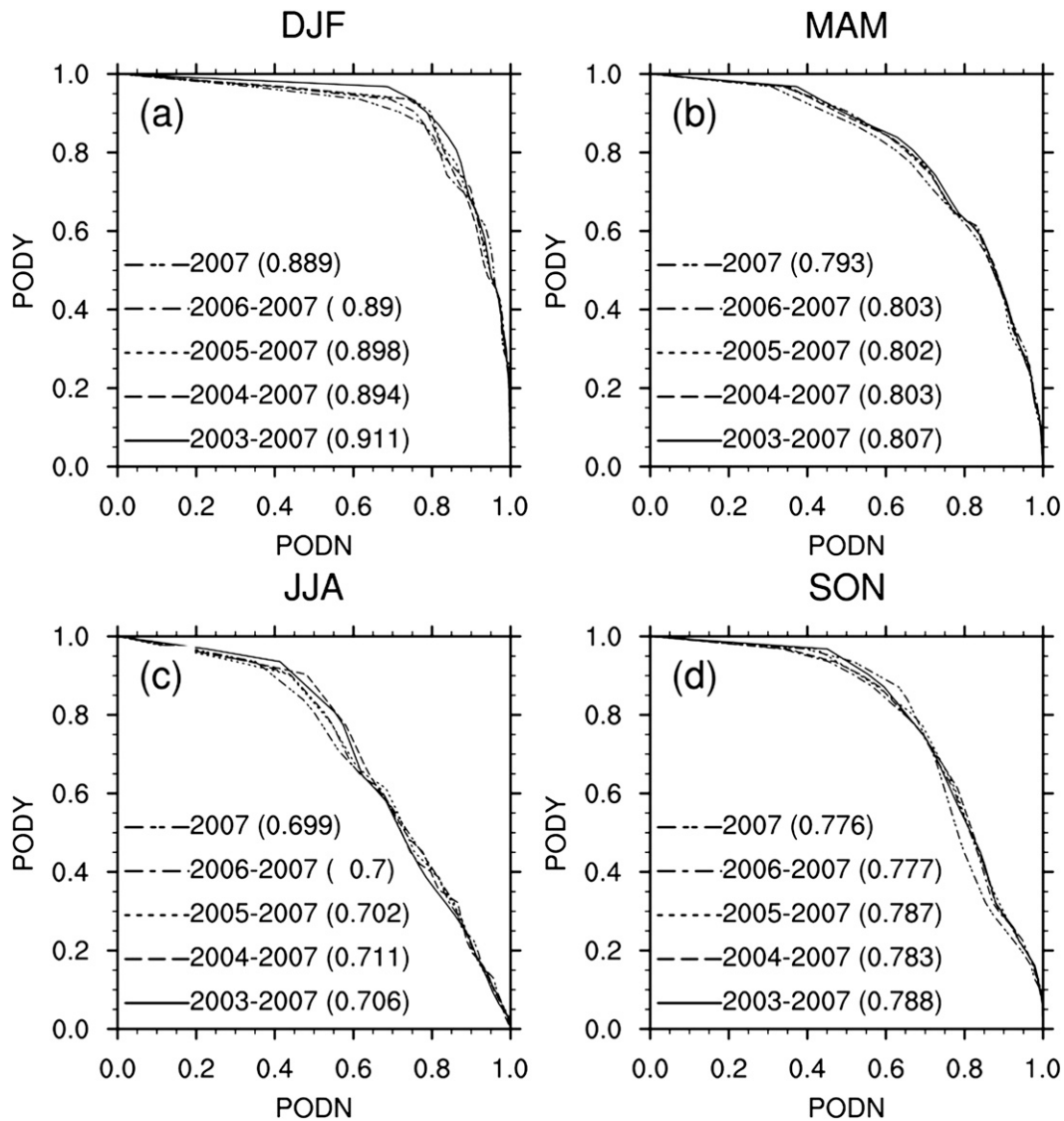


FIG. 7. PODY-PODN statistics of the 2008 GTG_{EA} for (a) DJF, (b) MAM, (c) JJA, and (d) SON that are based on the previous 1-yr (2007), 2-yr (2006-07), 3-yr (2005-07), 4-yr (2004-07), and 5-yr (2003-07) PIREP and RDAPS analysis data over East Asia. The AUC values for each GTG_{EA} combination are written in parentheses.

For the seasonal GTG_{EA} (Fig. 5b), the AUC values of the wintertime GTG_{EA} are higher than other seasonal experiments, whereas summertime AUC values are lower than the other seasons. The overall patterns of the seasonal AUC curves becomes higher as the number of combined indices increases up to 13-19; maximum values of AUC for December-February (DJF), March-May (MAM), June-August (JJA), and September-November (SON) GTG_{EA} are 0.852, 0.818, 0.795, and 0.821 using 19, 15, 13, and 15 of combined indices, respectively. Thus the GTG_{EA} performance over East Asia is optimal when approximately 15 component diagnostics are used.

When the performance of the yearly and seasonal GTG_{EA} , with weights and thresholds that are based on the yearly and seasonally classified PIREP data, is evaluated over East Asia (not shown), the results are not significantly different from those that are based on the 6-yr PIREP data (shown in Figs. 4 and 5) because of the similarity of the selected thresholds and weighting combinations that are based on the 6-yr PIREP data (Table 1) and the yearly and seasonally classified PIREP data (not shown).

Next, the sensitivity of the AUC performance metric to the amount of historical PIREP data that is available is investigated. The PODY-PODN statistics of the 2008 GTG_{EA} obtained from the 20 diagnostics that are based

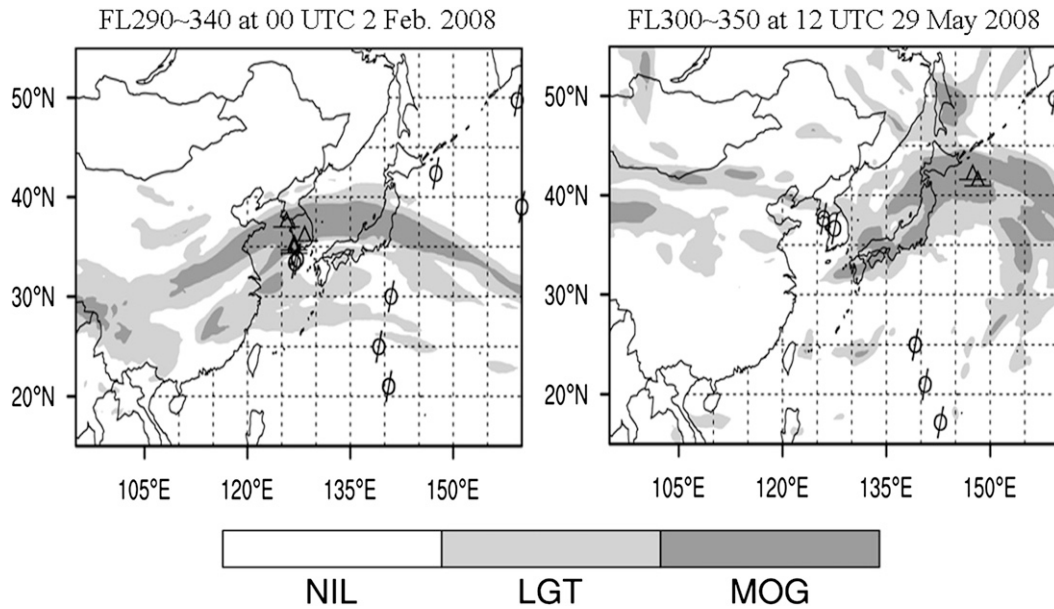


FIG. 8. As in Fig. 3, but on the GFS domain focused on the East Asia region, derived using the selected 20 diagnostics and thresholds listed in Table 3, on the basis of the 1-yr (2008) PIREP data.

on the previous 1-yr (2007), 2-yr (2006–07), 3-yr (2005–07), 4-yr (2004–07), and 5-yr (2003–07) PIREP data are depicted in Fig. 6. Although the performances of the 2008 GTG_{EA} that are based on the previous years' PIREP data are similar to each other, the skill becomes better as the number of PIREP samples increases from 1 year (2007) to 5 years (2003–07).

For the seasonal GTG_{EA} performance, the DJF, MAM, JJA, and SON performances of the 2008 GTG_{EA} obtained from the selected 20 best diagnostics that are based on the previous 1-yr (2007), 2-yr (2006–07), 3-yr (2005–07), 4-yr (2004–07), and 5-yr (2003–07) seasonal PIREP data are depicted in Fig. 7. The best performance of the wintertime forecasting experiments for the 2008 GTG_{EA} is 0.911, which uses the selected 20 diagnostics and weightings of the GTG_{EA} that are based on the previous 5-yr (2003–07) wintertime PIREP data (Fig. 7a). In general, the seasonal forecasting skill of the 2008 GTG_{EA} becomes higher as the number of seasonal PIREP samples increases from the previous 1 year (2007) to 5 years (2003–07), although the increases are not significant in all seasonal GTG_{EA} calculations.

b. Comparison between the RDAPS- and GFS-based GTG_{EA}

In this section the GFS analysis data with a 0.3125° horizontal grid spacing (T382) are used to calculate the GTG_{EA} , and the effects of the NWP model output on the GTG_{EA} performance are investigated by comparing the results with those presented previously using the

RDAPS analysis data. If the large-scale atmospheric flows, and the horizontal and vertical gradients associated with those flows, are more accurately represented by the underlying NWP model, the GTG_{EA} diagnoses and forecasts should improve. The GFS analysis data used in this study are provided on 37 isobaric surfaces from the 1000- to 100-hPa levels and are focused on the East Asia region (Fig. 8). The difference in the horizontal resolution over the East Asia region between two NWP models is not significant, but the vertical resolution in the GFS is higher. The GFS data with a 0.3125° horizontal grid spacing have been archived at NCAR since November of 2007, and therefore the evaluations and comparisons of the GTG system in this section are conducted using the 1-yr (2008; December 2007–November 2008) PIREP data within ± 2 h around 0000 and 1200 UTC. Hereinafter, the 2008 GTG_{EA} derived using the RDAPS and GFS data on the basis of the 2008 PIREP data will be indicated as the RDAPS-12 and GFS-12 experiments, respectively. The GTG_{EA} procedures used in the GFS-12 are the same as those in the RDAPS-12.

The first and second columns of Table 3 list the selected diagnostics and their AUC performance in the RDAPS-12 and GFS-12 experiments. Two interesting features are found in Table 3. First, the performance of the individual diagnostics, as well as the 2008 GTG_{EA} , is better in the GFS-12 experiment than in the RDAPS-12 experiment. This is due in large part to the better performance of the “Ri” diagnostic that is based on the

TABLE 3. Index name and AUC values of the 2008 GTG_{EA} combinations for the RDAPS-12, GFS-12, and GFS-06 experiments and their 20 component diagnostics that are based on the 1-yr (2008) PIREP data. The numbers of the MOG- and NIL-level events for each experiment are written in parentheses. Diagnostics selected for all experiments are shown in boldface.

RDAPS-12 (264; 4377)		GFS-12 (264; 4377)		GFS-06 (554; 9759)	
Index	AUC	Index	AUC	Index	AUC
GTG	0.787	GTG	0.823	GTG	0.831
DTF3	0.791	-Ri_e	0.815	-Ri_e	0.830
CP	0.754	DTF3	0.795	DTF3	0.826
-Ri_e	0.750	CP	0.790	CP	0.807
Brown2	0.744	Brown2	0.759	EDRLT	0.763
F_θ	0.737	TH1	0.757	TH1	0.751
TH1	0.735	EDRTL	0.751	Brown2	0.748
VWS	0.717	EDRLT	0.743	EDRTL	0.743
EDRLL	0.712	EDRLL	0.736	Dutton	0.732
EDRLT	0.682	Dutton	0.730	EDRLL	0.729
Dutton	0.679	VWS	0.721	VWS	0.719
TROPV	0.679	DEF	0.694	NGM2	0.714
HTG	0.677	Endlich	0.688	F_θ	0.703
Endlich	0.674	w	0.675	w	0.700
EDRTL	0.657	DIV	0.671	DEF	0.688
NVA	0.652	F_θ	0.671	Endlich	0.678
CCAT	0.650	NGM2	0.665	IAwind	0.664
DEF	0.644	HTG	0.663	DIV	0.652
HPG	0.629	HPG	0.653	HTG	0.639
DIV	0.625	VORT	0.637	LHF	0.624
NGM2	0.624	IAwind	0.626	VORT	0.624

GFS, which is presumably due to the better vertical resolution. Second, 17 of 20 best-performing diagnostics are common to both the RDAPS-12 and GFS-12 experiments, although the performance of these diagnostics varies. This implies that the performance of the 2008 GTG_{EA} over East Asia depends on the NWP model used, and the performance becomes better as the vertical resolution of the NWP model increases, although note that the improvements in the GFS-based GTG_{EA} could also be caused by differences in the initialization procedures and physical packages in the NWP model. Figure 8 shows examples of the 2008 GTG_{EA} in the GFS-12 experiments at FL290–340 for 0000 UTC 2 February 2008 (left panel) and at FL300–350 for 1200 UTC 29 May 2008 (right panel), derived using the selected 20 diagnostics listed in Table 3. The locations of several PIREPs, including MOG- and NIL-level events that occurred within ± 2 h of the targeted times, are also displayed. The 2008 GTG_{EA} in the GFS-12 experiment shown in Fig. 8 more correctly and discriminatively predicts the locations of all MOG- and NIL-level events over East Asia than does that in the RDAPS-12 experiment shown in Fig. 3.

To acquire a more detailed comparison between the RDAPS-12 and GFS-12, the PODY–PODN statistics

of the 2008 GTG_{EA} with the maximum and minimum boundaries of an additional 200 experiments for the RDAPS-12 and GFS-12, derived using 200 subsets of randomly selected half-fraction samples that are based on the 1-yr (2008) PIREP data, are shown in Figs. 9a and 9b. The PODY–PODN statistics of the 2008 seasonal GTG_{EA} from the RDAPS-12 and GFS-12 are also shown in Figs. 9c and 9d, respectively, to compare the seasonal dependencies of the RDAPS-based GTG_{EA} with those of the GFS-based GTG_{EA}. In Figs. 9a and 9b, the additional 200 experiments for 2008 and 2008 seasonal GTG_{EA} combinations are consistent with each other within about $\pm 5\%$ – 8% (RDAPS-12) and $\pm 4\%$ – 5% (GFS-12), implying that the improvement of the GTG_{EA} performance from the RDAPS-12 (0.787) to the GFS-12 (0.823) is statistically significant. In Figs. 9c and 9d, the performance values of the seasonal GTG_{EA} also increased from the RDAPS-12 to the GFS-12.

To determine the sensitivity to the number of combined diagnostics, the 2008 GTG_{EA} from the GFS-12 and RDAPS-12 experiments are depicted in Fig. 10 as a function of the number of indices used. The AUC values of the RDAPS-12 and GFS-12 increase as the number of component diagnostics increases up to 13 and 16, with optimal values for the RDAPS-12 and GFS-12 of 0.800 and 0.828, respectively. In general, the 2008 GTG_{EA} that is based on the GFS system, which has finer vertical grid spacing, outperforms the 2008 GTG_{EA} that is based on the RDAPS analysis data.

c. Sensitivity to temporal variations in the GFS and PIREP data

The GFS analysis data are available 4 times per day at 0000, 0600, 1200, and 1800 UTC, whereas the RDAPS analysis data are only produced twice daily at 0000 and 1200 UTC. In the previous section, only the GFS data at 0000 and 1200 UTC (denoted GFS-12) were used to directly compare with the GTG_{EA} calculations using the RDAPS analysis. In this section 6-hourly GFS data, with corresponding PIREPs that occurred within ± 2 h, are used to provide more data to evaluate the GFS-based GTG_{EA} performance. The third column of Table 3 shows the AUC performance of the 2008 GTG_{EA} obtained from the selected 20 diagnostics that are based on the 1-yr (2008) PIREP recorded data within ± 2 h of 0000, 0600, 1200, and 1800 UTC. Hereinafter, this experiment is denoted as GFS-06.

The numbers of the NIL- and MOG-level events for the GFS-06 experiment are 9759 and 554, respectively, which are more than 2 times those (4377 and 264) for the GFS-12 experiment. The comparison between the

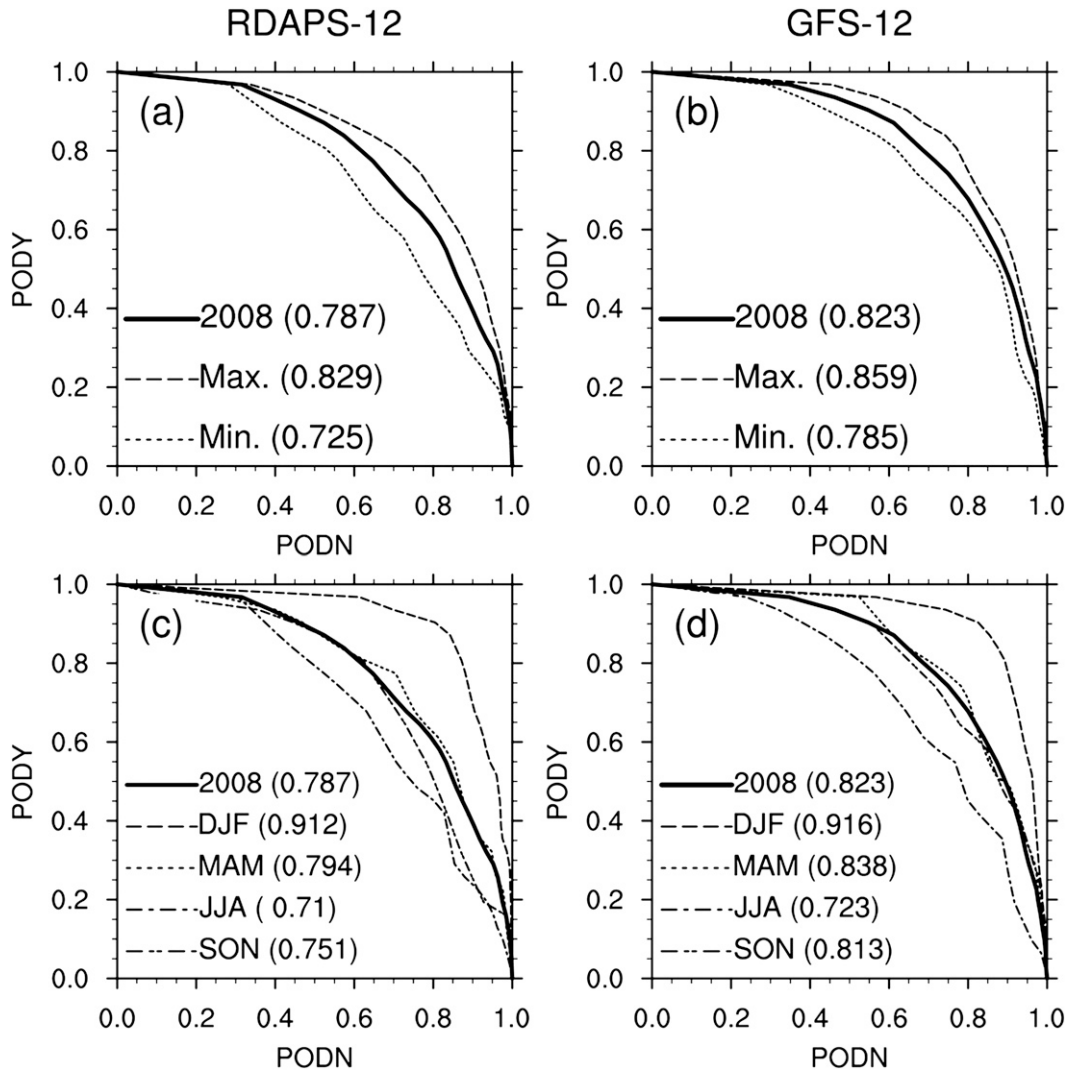


FIG. 9. PODY–PODN statistics of the 2008 GTG_{EA} derived from the (left) RDAPS-12 and (right) GFS-12 experiments with (a),(b) the maximum and minimum boundaries of the additional 200 experiments using 200 subsets of randomly selected half fraction samples and (c),(d) those for each season that are based on the 1-yr (2008) PIREP data over East Asia. In (c) and (d), plots of DJF, MAM, JJA, and SON are provided. The AUC values for each GTG_{EA} combination are written in parentheses.

GFS-12 and GFS-06 experiments in Table 3 showed two interesting features. First, the performance of most of the component diagnostics as well as the 2008 GTG_{EA} in the GFS-06 experiment is higher than those in the GFS-12 experiment. Second, 19 of 20 diagnostics are common to both the GFS-06 and GFS-12 experiments (see Table 3), although the combinations and weightings of the selected 20 diagnostics are slightly different from each other.

To investigate the sensitivity of the 2008 GTG_{EA} GFS-06 experiment to the number of the indices used, AUCs are depicted in Fig. 10 as a function of the number of combined indices. In general, the GTG_{EA} skill in

the GFS-06 experiment becomes higher as the number of combined indices increases up to 15 (AUC = 0.838) but decreases or becomes flat as still more indices are included. In comparing the GFS-06 AUCs with the GFS-12 AUCs in Fig. 10, it is seen that the GFS-06 experiments have higher AUCs than the GFS-12 in all combinations with 2–20 component diagnostics. This implies that the skill of the 2008 GTG_{EA} over East Asia increases as the finer temporal resolution of the NWP data is used. As a consequence, more PIREP data that reduce diurnal dependency in the PIREP data (shown in Fig. 1) are applied to the determination of the GTG_{EA} diagnostic weights. This is consistent with previous

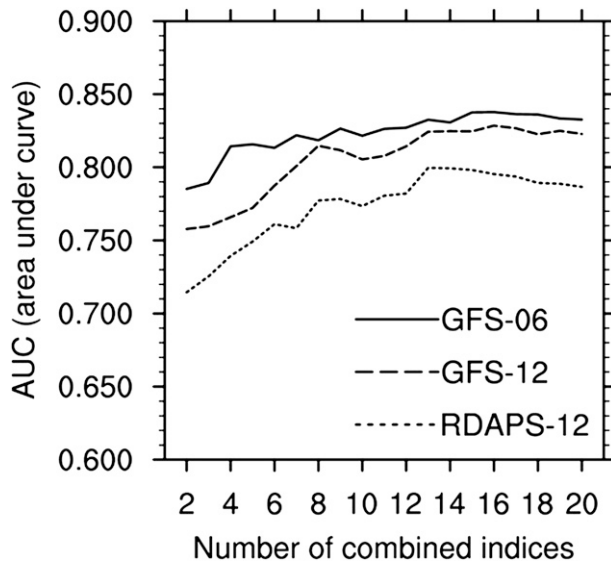


FIG. 10. AUCs of the 2008 GTG_{EA} as a function of the number of combined indices, derived from the RDAPS-12, GFS-12, and GFS-06 experiments, that are based on the 1-yr (2008) PIREP data over East Asia.

studies in that using more PIREP data to determine diagnostic weights improves the GTG performance (Sharman et al. 2006; Kim et al. 2009).

d. Sensitivity to the forecast lead time of the GFS-based GTG_{EA}

In this section, the 2008 GTG_{EA} combination from the GFS-06 experiment listed in the third column of Table 3 is applied to the 6- and 12-h forecast GFS data to evaluate the GTG_{EA} performance as a function of forecast lead time. Figure 11 shows the PODY-PODN statistics of the 2008 GTG_{EA} that are based on the 1-yr (2008) PIREP data and the GFS analyses (GFS-06), 6-h forecast (FCST06), and 12-h forecast (FCST12). In addition, experiments for the GFS-06, FCST06, and FCST12 are conducted using 200 subsets of randomly selected half-fraction samples for each experiment to investigate whether the results are statistically significant. Interesting features found in Fig. 11 are as follows. First, the 2008 GTG_{EA} performance decreased from 0.831 to 0.819 as the forecast lead time increased from 0 to 12 h (Fig. 11a). This is somewhat expected, but the magnitudes of the decrease is not as large as that in Sharman et al. (2006), which showed that the GTGC AUC for upper levels are 0.878 (analyses) and 0.852 (6-h forecasts). Second, the uncertainty boundaries of the GFS-06 (Fig. 11b), FCST06 (Fig. 11c), and FCST12 (Fig. 11d) experiments are about $\pm 3\%$ – 5% , implying that the demonstrated decrease in GTG_{EA} performance

with increasing forecast lead time is statistically significant.

To examine the dependency in the number of component diagnostics along with the forecast lead time of the GFS-06 experiment, AUCs of the 2008 GTG_{EA} as a function of the number of combined indices from the GFS-06, FCST06, and FCST12 experiments are shown in Fig. 12. As can be seen in Fig. 12, the 2008 GTG_{EA} performance with 2–20 component diagnostics always decreased as the forecast lead time increased from 0 to 12 h. The optimal AUC values in the GFS-06, FCST06, and FCST12 experiments are 0.838, 0.833, and 0.822 with 15, 16, and 17 component diagnostics, respectively.

5. Summary and conclusions

The performance of upper-level turbulence diagnostics over East Asia using a GTG system is investigated using various combinations of diagnostics derived from two different underlying operational NWP models. The GTG system combines a suite of turbulence diagnostics to obtain the turbulence potential, and although the GTG system has been used in the United States since 2003 [on the basis of the Rapid Update Cycle 13 (RUC13) NWP model] and seems to provide good statistical performance (Sharman et al. 2006), its use in other environments and with other operational NWP models requires careful evaluation. Here, this evaluation has been done over East Asia using six years of PIREPs for verification. Eight key findings came out of this study:

- 1) The forecast performance (AUC, TSS, and RMSE) of the optimal GTG_{EA} combination is always superior to that of the single best diagnostic.
- 2) The optimal suite of diagnostics found in this study is somewhat different from the set used in the Sharman et al. (2006) GTG system, indicating an environmental and NWP-model dependence on algorithm performance. This is likely because the jet stream over East Asia is much stronger climatologically than it is over the CONUS and the spatiotemporal variations of the jet stream in the two regions are different. Given that the shear and inertial instabilities associated with the jet stream are the major generation mechanisms of observed turbulence encounters, this would be expected to have a major impact on the skill of the individual diagnostics.
- 3) The GTG_{EA} performance obtained from the selected 20 best diagnostics depends on year and season, because of the annual and seasonal variations of the large-scale atmospheric flows that can affect aircraft-scale turbulence over East Asia. The annual

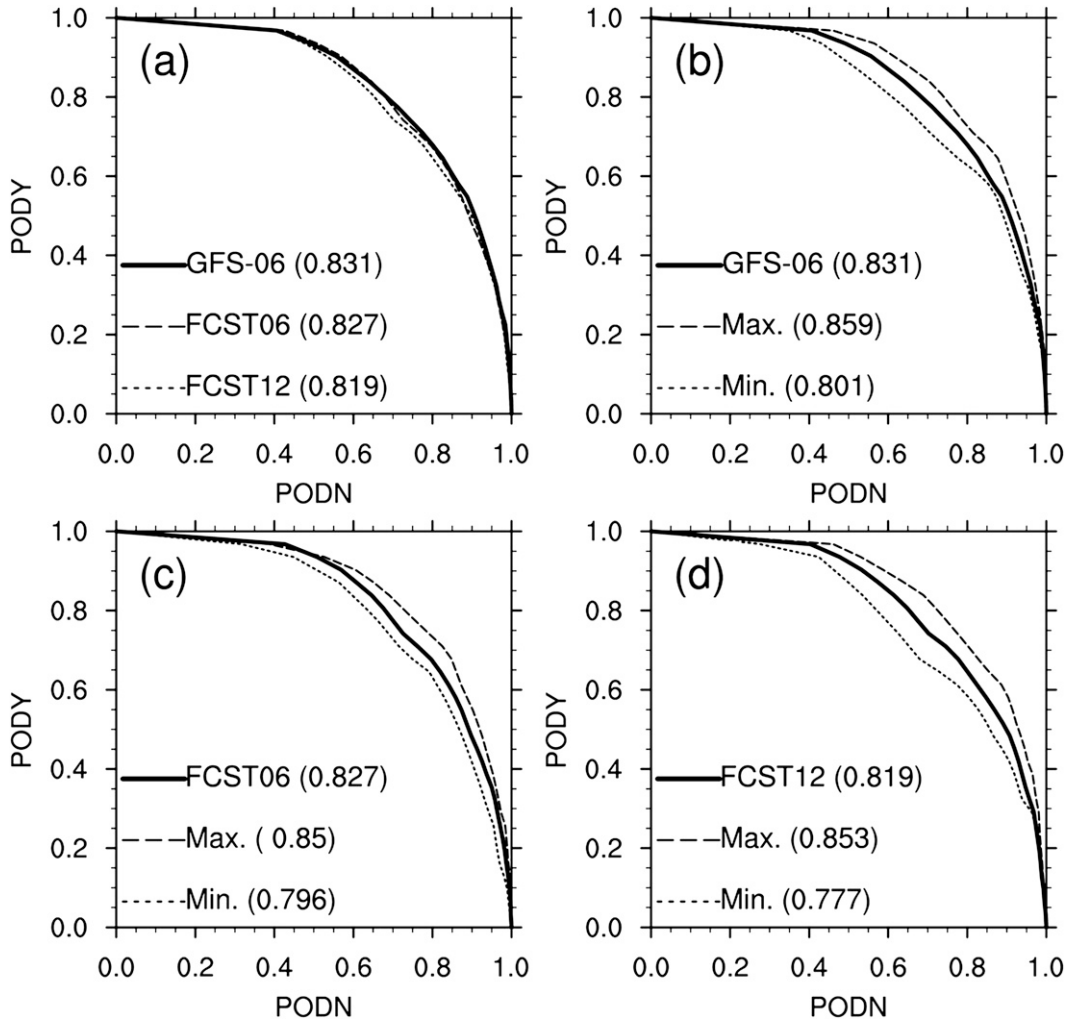


FIG. 11. (a) PODY–PODN statistics of the 2008 GTG_{EA} of the GFS-06 experiment, derived from the analyses (GFS-06), 6-h forecasts (FCST06), and 12-h forecasts (FCST12) data and the 1-yr (2008) PIREP data over East Asia. PODY–PODN statistics of the (b) GFS-06, (c) FCST06, and (d) FCST12 experiments with the maximum and minimum boundaries of 200 experiments using subsets of randomly selected half-fraction samples. The AUC values for each GTG_{EA} combination are written in parentheses.

and seasonal dependencies in the GTG_{EA} performance are statistically significant.

- 4) The wintertime GTG_{EA} skill has the best seasonal performance. This is likely because the current GTG system includes diagnostics that are heavily weighted to jet stream and upper-level frontogenesis processes that are most pronounced during the wintertime. The summertime GTG_{EA} skill is much lower than for other seasons. This implies that new turbulence diagnostics that properly detect turbulence events during the summertime (which are probably related to convection) should be developed and implemented in future GTG systems.
- 5) Performance increases as the number of turbulence diagnostics used in the GTG_{EA} combinations

increases up to about 15 diagnostics, after which there is no noticeable improvement.

- 6) In general, the yearly GTG_{EA} performance increases with the number of PIREPs used in the development of the individual diagnostic weights and thresholds. Thus using the previous 5 yr of PIREP data to tune GTG_{EA} for year 6 (2008 GTG_{EA}) gave the best performance.
- 7) Finer vertical resolution of the underlying NWP model seems to provide consequent improvements in GTG_{EA} skill. Note, however, that the improvements of the GTG_{EA} skill could be caused not only by the increase of the vertical resolution but also by the initialization procedures and physical parameterizations that are used in the NWP model.

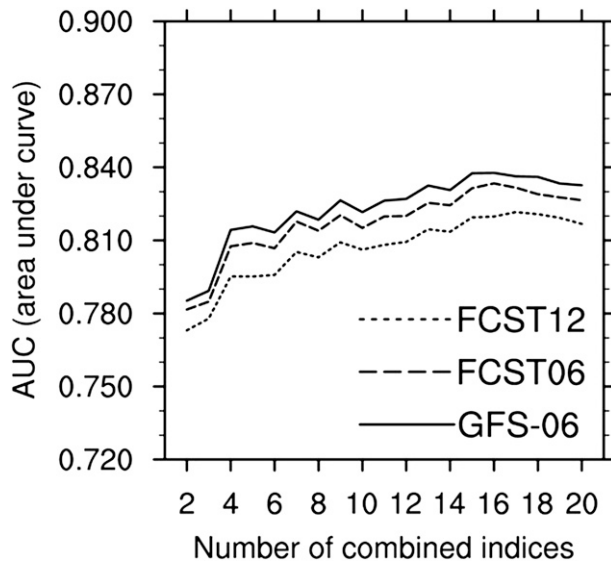


FIG. 12. AUCs of the 2008 GTG_{EA} as a function of the number of combined indices, derived from the GFS-06 experiment using analyses (GFS-06), 6-h forecasts (FCST06), and 12-h forecasts (FCST12), that are based on the 1-yr (2008) PIREP data over East Asia.

- 8) Forecasting performance of the 2008 GTG_{EA} becomes lower as the forecasting lead time increases from 0 to 12 h.

Although a few difficulties, already stated in Sharman et al. (2006), still remain in the current study, these evaluations can provide useful information for the development of an upper-level turbulence forecasting GTG system over East Asia. These difficulties are 1) the coarse resolution of the current NWP model leading to an inability to resolve aircraft-scale turbulence, 2) the lack of understanding of the linkage between NWP-resolvable-scale flows and aircraft-scale turbulence, and 3) the forecast errors in upper-level flows in the NWP model.

The optimal compositions of the GTG system found in this study, such as the way to select the turbulence diagnostics (i.e., using yearly or seasonal PIREP data) and the number of diagnostics, should be tested in areas other than the East Asia region as well. To be specific, because of the lack of PIREPs over China, the GTG_{EA} performance over China using the setup derived here may not be optimal.

Note last that the GTG combination, even though it is based on the computation of 20 different diagnostics, is not computationally intensive. Using the climatological thresholds ($GTGC$) and weightings for the individual diagnostics used in the GTG system (Table 1), the GTG output is available within about 10 minutes, at least in the current calculation domain near the East Asia, once the operational forecasting model data are available.

Acknowledgments. This work was funded by the Korean Meteorological Administration Research and Development Program under Grant RACS_2011-8006.

REFERENCES

- Abernethy, J. A., 2008: A domain analysis approach to clear-air turbulence forecasting using high-density in-situ measurements. Ph.D. Dissertation, Department of Computer Science, University of Colorado, 115 pp.
- Brown, B. G., J. L. Mahoney, J. Henderson, T. L. Kane, R. Bullock, and J. E. Hart, 2000: The turbulence algorithm intercomparison exercise: Statistical verification results. Preprints, *Ninth Conf. on Aviation, Range, and Aerospace Meteorology*, Orlando, FL, Amer. Meteor. Soc., 7.4. [Available online at http://ams.confex.com/ams/Sept2000/techprogram/paper_16470.htm.]
- Brown, R., 1973: New indices to locate clear-air turbulence. *Meteor. Mag.*, **102**, 347–360.
- Cho, J. Y. N., and E. Lindborg, 2001: Horizontal velocity structure functions in the upper troposphere and lower stratosphere 1. Observations. *J. Geophys. Res.*, **106**, 10 223–10 232.
- Clark, T. L., and Coauthors, 2000: Origins of aircraft-damaging clear air turbulence during the 9 December 1992 Colorado downslope windstorm: Numerical simulations and comparison to observations. *J. Atmos. Sci.*, **57**, 1105–1131.
- Colson, D., and H. A. Panofsky, 1965: An index of clear-air turbulence. *Quart. J. Roy. Meteor. Soc.*, **91**, 507–513.
- Dutton, J. A., and H. A. Panofsky, 1970: Clear air turbulence: A mystery may be unfolding. *Science*, **167**, 937–944.
- Ellrod, G., and D. L. Knapp, 1992: An objective clear-air turbulence forecasting technique: Verification and operational use. *Wea. Forecasting*, **7**, 150–165.
- , and J. Knox, 2010: Improvements to an operational clear-air turbulence diagnostic index by addition of a divergence trend term. *Wea. Forecasting*, **25**, 789–798.
- , P. F. Lester, and L. J. Ehernberger, 2003: Clear air turbulence. *Encyclopedia of Atmospheric Sciences*, J. R. Holton et al., Eds., Vol. 1, Academic Press, 393–403.
- Feltz, W. F., K. M. Bedka, J. A. Otkin, T. Greenward, and S. A. Ackerman, 2009: Understanding satellite-observed mountain-wave signatures using high-resolution numerical model data. *Wea. Forecasting*, **24**, 76–86.
- 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.*, **132**, 2308–2324.
- Jaeger, E. B., and M. Sprenger, 2007: A Northern Hemispheric climatology of indices for clear air turbulence in the tropopause region derived from ERA40 reanalysis data. *J. Geophys. Res.*, **112**, D20106, doi:10.1029/2006JD008189.
- Jang, W., H.-Y. Chun, and J.-H. Kim, 2009: A study of forecast system for clear-air turbulence in Korea. Part I: Korean Integrated Turbulence Forecasting Algorithm (KITFA) (in Korean with English abstract). *Atmosphere (Toronto)*, **19** (3), 255–268.
- Kaplan, M., and D. R. Vollmer, 2010: The meteorological environment surrounding the Air France #447 disaster. Preprints, *14th Conf. on Aviation Range and Aerospace Meteorology*, Atlanta, GA, Amer. Meteor. Soc., 11.1. [Available online at <http://ams.confex.com/ams/pdfpapers/160191.pdf>.]
- Kim, J.-H., and H.-Y. Chun, 2010: A numerical study of clear-air turbulence (CAT) encounters over South Korea on 2 April 2007. *J. Appl. Meteor. Climatol.*, **49**, 2381–2403.

- , and —, 2011: Statistics and possible sources of aviation turbulence over South Korea. *J. Appl. Meteor. Climatol.*, **50**, 311–324.
- , —, W. Jang, and R. D. Sharman, 2009: A study of forecast system for clear-air turbulence in Korea. Part II: Graphical turbulence guidance (GTG) system (in Korean with English abstract). *Atmosphere (Toronto)*, **19** (3), 269–287.
- Knox, J. A., D. W. McCann, and P. D. Williams, 2008: Application of the Lighthill–Ford theory of spontaneous imbalance to clear-air turbulence forecasting. *J. Atmos. Sci.*, **65**, 3292–3304.
- , A. S. Bachmeier, W. M. Carter, J. E. Tarantino, L. C. Paulik, E. N. Wilson, G. S. Bechdol, and M. J. Mays, 2010: Transverse cirrus bands in weather systems: A grand tour of an enduring enigma. *Weather*, **65** (2), 35–41.
- Koch, P., H. Wernli, and H. W. Davies, 2006: An event-based jet-stream climatology and typology. *Int. J. Climatol.*, **26**, 283–301.
- Koshyk, J. N., and K. Hamilton, 2001: The horizontal energy spectrum and spectral budget simulated by a high-resolution troposphere–stratosphere–mesosphere GCM. *J. Atmos. Sci.*, **58**, 329–348.
- Lane, T. P., R. D. Sharman, T. L. Clark, and H.-M. Hsu, 2003: An investigation of turbulence generation mechanisms above deep convection. *J. Atmos. Sci.*, **60**, 1297–1321.
- , J. D. Doyle, R. D. Sharman, M. A. Shapiro, and C. D. Watson, 2009: Statistics and dynamics of aircraft encounters of turbulence over Greenland. *Mon. Wea. Rev.*, **137**, 2687–2702.
- Lee, Y.-G., B.-C. Choi, R. Sharman, G. Wiener, and H.-W. Lee, 2003: Determination of the primary diagnostics for the CAT (clear-air turbulence) forecast in Korea. *J. Korean Meteor. Soc.*, **39**, 677–688.
- Lenz, A., K. M. Bedka, W. F. Feltz, and S. A. Ackerman, 2009: Convectively induced transverse band signatures in satellite imagery. *Wea. Forecasting*, **24**, 1362–1373.
- Lester, P. F., 1994: *Turbulence: A New Perspective for Pilots*. Jeppesen Sanderson, 212 pp.
- Marroquin, A., 1998: An advanced algorithm to diagnose atmospheric turbulence using numerical model output. Preprints, *16th Conf. on Weather Analysis and Forecasting*, Phoenix, AZ, Amer. Meteor. Soc., 79–81.
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 677–688.
- National Transportation Safety Board, 2009: U.S. Air Carrier Operations, Calendar Year 2005. Annual review of aircraft accident data. NTSB/ARC-09/01, Washington, DC, 66 pp.
- Olafsson, H., and H. Agustsson, 2009: Gravity wave breaking in easterly flow over Greenland and associated low level barrier and reverse tip-jets. *Meteor. Atmos. Phys.*, **104**, 191–197.
- 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.
- Tebaldi, C., D. Nychka, B. G. Brown, and R. Sharman, 2002: Flexible discriminant techniques for forecasting clear-air turbulence. *Environmetrics*, **13**, 859–878.
- Tung, K. K., and W. W. Orlando, 2003: The k^{-3} and $k^{-5/3}$ energy spectrum of atmospheric turbulence: Quasigeostrophic two-level model simulation. *J. Atmos. Sci.*, **60**, 824–835.
- Uhlenbrock, N. L., K. M. Bedka, W. F. Feltz, and S. A. Ackerman, 2007: Mountain wave signatures in MODIS 6.7- μm imagery and their relation to pilot reports of turbulence. *Wea. Forecasting*, **22**, 662–670.
- Williams, J. K., C. J. Kessinger, R. D. Sharman, W. F. Feltz, and A. Wimmers, 2010: A probabilistic global turbulence nowcast and forecast system. Preprints, *14th Conf. on Aviation Range and Aerospace Meteorology*, Atlanta, GA, Amer. Meteor. Soc., J11.7. [Available online at http://ams.confex.com/ams/90annual/techprogram/paper_164700.htm.]
- Wolff, J., and R. Sharman, 2008: Climatology of upper-level turbulence over the continental United States. *J. Appl. Meteor. Climatol.*, **47**, 2198–2214.

CORRIGENDUM

JUNG-HOON KIM AND HYE-YEONG CHUN

Department of Atmospheric Sciences, Yonsei University, Seoul, South Korea

ROBERT D. SHARMAN AND TEDDIE L. KELLER

Research Application Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

Because of a production error, the journal titles of two references in Kim et al. (2011) were listed incorrectly. The modifier “(Toronto)” was incorrectly inserted after the correct journal title “*Atmosphere*.” The corrected reference entries for Jang et al. (2009) and Kim et al. (2009) are given below in the reference section. The staff of the *Journal of Applied Meteorology and Climatology* regrets any inconvenience this error may have caused.

REFERENCES

- Jang, W., H.-Y. Chun, and J.-H. Kim, 2009: A study of forecast system for clear-air turbulence in Korea. Part I: Korean Integrated Turbulence Forecasting Algorithm (KITFA) (in Korean with English abstract). *Atmosphere*, **19** (3), 255–268.
- Kim, J.-H., H.-Y. Chun, W. Jang, and R. D. Sharman, 2009: A study of forecast system for clear-air turbulence in Korea. Part II: Graphical Turbulence Guidance (GTG) system (in Korean with English abstract). *Atmosphere*, **19** (3), 269–287.
- , —, R. D. Sharman, and T. L. Keller, 2011: Evaluations of upper-level turbulence diagnostics performance using the Graphical Turbulence Guidance (GTG) system and pilot reports (PIREPs) over East Asia. *J. Appl. Meteor. Climatol.*, **50**, 1936–1951.

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Prof. Hye-Yeong Chun, Dept. of Atmospheric Sciences, Yonsei University, 262 Seongsanno, Seodaemun-gu, Seoul 120-749, South Korea.
E-mail: chunhy@yonsei.ac.kr