Evaluation of Multimodel-Based Ensemble Forecasts for Clear-Air Turbulence[✉]

DAN-BI LEE AND HYE-YEONG CHUN

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

JUNG-HOON KIM

School of Earth and Environmental Sciences, Seoul National University, Seoul, South Korea

(Manuscript received 23 July 2019, in final form 23 January 2020)

ABSTRACT

To test more consistent and reliable upper-level turbulence forecasts, seven global numerical weather prediction (NWP) model outputs are used to construct the multimodel-based ensemble forecasts for clear-air turbulence (CAT). We used the updated version of the well-known Ellrod index, the Ellrod–Knox index (EKI), which is currently an operational CAT diagnostic for the significant weather chart at one of the World Area Forecast Centers. In this study, we tested two types of ensemble forecasts. First is an ensemble mean of all EKI forecasts from the NWP models. Second is a probabilistic forecast that is computed by counting how many individual EKI values from the seven NWP models exceed a certain EKI threshold at each grid point. Here, to calibrate the best EKI thresholds for the moderate-or-greater CAT intensity, the individual EKI thresholds, which vary depending on the resolutions and configurations of the NWP models, are selected using the 95th, 98th, and 99th percentiles of the probability density functions for the EKIs derived from the seven NWP models for a 6-month period. Finally, performance skills of both the ensemble mean and probabilistic forecasts are evaluated against the observations of in situ aircraft eddy dissipation rate and pilot reports. As a result, the ensemble mean forecast shows a better performance skill than the individual EKI forecasts. The reliability diagram for the probabilistic forecast gives a better reliability when using high-percentile EKI values as the threshold although it still suffers overestimation of CAT events likely due to the lack of observation and ensemble spreads.

1. Introduction

Clear-air turbulence (CAT) is small-scale turbulent eddies in the free atmosphere that directly affect cruising aircraft usually in clear sky conditions (Sharman and Lane 2016). CAT frequently occurs in the vicinity of upper-level jet and frontal systems in which strong vertical wind shear (VWS) leads to Kelvin–Helmholtz instability, which can be quantified by the Richardson number (Ri) (e.g., Dutton and Panofsky 1970; Kim and Chun 2010). Unbalanced flows in the exit region of the jet stream are associated with the emission of the inertia–gravity waves (IGWs) (Uccellini and Koch 1987; Guest et al. 2000; Plougonven et al. 2003). This eventually leads

to activating CAT through the reduction of Ri (Knox et al. 2008; Kim and Chun 2011; Sharman et al. 2012). Anticyclonic shear and curvature flows of upper-level ridges are also conducive to CAT production via geostrophic adjustment and inertial instability (Knox 1997; Kim and Chun 2010; Sharman et al. 2012).

Various CAT diagnostics have been developed for decades in aviation turbulence forecast systems using numerical weather prediction (NWP) models based on the concept of energy cascading down from resolvable synoptic scale or mesoscale to smaller scales (e.g., Dutton and Panofsky 1970; Lee et al. 1984; Ellrod and Knapp 1992; Kaplan et al. 2005a,b, 2006; Sharman et al. 2006; Ellrod and Knox 2010). Although the horizontal and vertical scales of CAT are still too small (~ 100 m) to be resolved directly from the current NWP models, CAT diagnostics derived by NWP models can provide potential locations of turbulence in the free atmosphere. Among various CAT diagnostics, the turbulence index 1 or Ellrod1 index (Ellrod and Knapp 1992), which combines VWS

[✉] Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/WAF-D-19-0155.s1>.

Corresponding author: Prof. Hye-Yeong Chun, chunhy@yonsei.ac.kr

and total deformation (DEF) to detect CAT in situations of upper-level frontogenesis, has been widely used among operational meteorological data providers worldwide for supporting strategic flight planning and preflight briefings for worldwide pilots and civil aviation users (Ellrod and Knapp 1992; Kim and Chun 2012; Sharman and Pearson 2017; Kim et al. 2018; Storer et al. 2019). Recently, Ellrod and Knox (2010) proposed an improved Ellrod1 index, called the Ellrod–Knox index (EKI), by adding a divergence tendency (DVT) term to account for CAT related to IGWs both in cyclonic and anticyclonic flows, which has been demonstrated as the highest performing diagnostic among various CAT diagnostics (Sharman and Pearson 2017). In this study, we choose EKI as a representative CAT diagnostic for multimodel-based ensemble forecasts for CAT.

To provide reliable probabilities for CAT forecasts, it is necessary that the CAT diagnostics need to take into account inherent uncertainties. To conciliate the uncertainties, probabilistic turbulence forecasts have been suggested in some previous studies using different ensemble system. Gill and Buchanan (2014) suggested a probabilistic turbulence forecast based on the Met Office Global and Regional Ensemble Prediction System (MOGREPS) forecasts with 24 ensemble members. For each turbulence diagnostic, a probabilistic turbulence forecast is produced by calculating a ratio of ensemble members that exceed a certain diagnostic threshold among the total ensemble members. They showed that the overall performance of the probabilistic forecast is better than that of the current operational deterministic forecasts from the World Area Forecast Centers (WAFCs). Kim et al. (2018) proposed a probabilistic aviation turbulence forecast using multiple turbulence diagnostics calculated using a deterministic forecast from the Global Ensemble Forecast System (GEFS). They demonstrated that the performance of the probabilistic forecast is better than that by the best individual turbulence diagnostic. As interests in multimodel ensembles have been increasing, several multimodel ensembles have been created based on either a group of ensemble forecasts of different systems or a group of deterministic forecasts of different systems (Bougeault et al. 2010). Storer et al. (2019) first applied the multimodel ensemble approach to the aviation turbulence forecast, based on two different ensemble models [European Centre for Medium-Range Weather Forecasts (ECMWF) and MOGREPS], and developed a multimodel ensemble-based probabilistic turbulence forecast.

In this study, the performance of EKI CAT forecasts based on the multimodel ensemble of seven individual deterministic forecasts from The International Grand Global Ensemble (TIGGE) database will be investigated.

To do this, we calculate two types of ensemble CAT forecasts. First is the ensemble mean of all EKI forecasts from the NWP models used. Second is a probabilistic CAT forecast using the EKI thresholds considering the resolution and configurations of the underlying NWP models. The results will be validated against archived turbulence observation data. The turbulence forecasting system constructed from this study can be extended to research for a super ensemble CAT forecast based on NWP ensemble forecasts as well as multiple diagnostics. Because the gradient of upper-level wind is tested and evaluated in various NWP models, this study can provide information on the model uncertainties related to upper-level jet/front dynamics, which can further assist model developers and data assimilation procedures. The remainder of this paper consists of three sections. In section 2, descriptions of the TIGGE data and EKI, along with the methodologies of the multimodel-based ensemble CAT forecasts will be given. In section 3, statistical evaluations of the CAT forecasts against global turbulence observation data will be described. A summary and conclusions will follow in section 4.

2. Methodology of multimodel-based ensemble forecasts for CAT

a. NWP model data

This study uses the global NWP model outputs from the TIGGE database to develop multimodel-based ensemble forecasts for CAT. TIGGE is a project that started in 2006 to support The Observing System Research and Predictability Experiment (THORPEX), which was established to accomplish rapid improvements in the accuracy of high-impact weather forecasts ranging from one day to two weeks (WMO 2005; Bougeault et al. 2010; Swinbank et al. 2016). The TIGGE database currently supplies the operational ensemble forecast data of 11 global forecast centers with a 2-day time delay via data archive portals at the ECMWF and China Meteorological Administration (CMA) for the purpose of scientific research (<http://apps.ecmwf.int/datasets/>). Because of the massive data volume, only selected parameters of temperature, geopotential height, specific humidity, and zonal and meridional winds at nine pressure levels (1000, 925, 850, 700, 500, 300, 250, 200, and 50 hPa) are used (Swinbank et al. 2016). TIGGE outputs are interpolated to a horizontal resolution of $0.5^\circ \times 0.5^\circ$, although the horizontal resolution of each original model is different. Note that even though all models are interpolated into the same horizontal resolution of 0.5° , the smooth (fine) features of meteorological variables in the

TABLE 1. Description of the TIGGE NWP models from the support forecast centers in 2017. Only 7 of the 11 TIGGE NWP models are used in this study to compute the MEM and MMP forecasts and are presented in the table.

Center	Horizontal resolution	Vertical level (model top)	Run time
China Meteorological Administration (CMA)	T213 (~70 km)	L31 (10 hPa)	0000, 1200 UTC
Brazil Centro de Previsão de Tempo Estudos Climático (CPTEC)	T126 (~120 km)	L28 (0.1 hPa)	0000, 1200 UTC
European Centre for Medium-Range Weather Forecasts (ECMWF)	TC639 (~18 km)	L91 (0.01 hPa)	0000, 1200 UTC
Japan Meteorological Agency (JMA)	T479 (~40 km)	L100 (0.01 hPa)	0000, 1200 UTC
Korea Meteorological Administration (KMA)	N400 (~32 km)	L70 (80 km)	0000, 1200 UTC
National Centers for Environmental Prediction (NCEP)	T574 (~34 km)	L64 (2.7 hPa)	0000, 0600, 1200, 1800 UTC
Met Office (UKMO)	N400 (~32 km)	L70 (80 km)	0000, 0600, 1200, 1800 UTC

underlying model with a coarse (fine) grid spacing remain in the interpolated model. The sensitivity of the underlying model resolution to the EKI forecast will be discussed in Figs. 2 and 3.

In this study, $T + 30$ h forecast is considered, which is a common time range for flight planning. Additionally, the largest volume of in situ turbulence observation data are recorded from the U.S. carriers during the daytime (around 1800 UTC) (Sharman and Pearson 2017; Kim et al. 2018). According to those reasons, this study focuses on multimodel-based CAT forecasts using 30-h control forecasts valid at 1800 UTC. For calculating the VWS term in EKI, data at 300-, 250-, and 200-hPa levels are used, which correspond to the cruising altitudes in the upper troposphere and lower stratosphere (UTLS) (Williams and Joshi 2013; Williams 2016; Kim et al. 2016). Here, only 7 of the 11 TIGGE model datasets are available to compute the multimodel-based ensemble forecasts for CAT during the evaluation period (October 2016–March 2017). A brief description of the seven TIGGE model datasets with the name of each center is shown in Table 1.

b. Computing the Ellrod–Knox index

In this study, EKI is used for the multimodel-based ensemble forecasts for CAT (Ellrod and Knox 2010) and consists of Ellrod1 and DVT as follows:

$$\text{EKI} = \text{Ellrod1} + \text{DVT}. \quad (1)$$

Here, Ellrod1 (Ellrod and Knapp 1992) and DVT are calculated by the following equations:

$$\text{Ellrod1} = \text{VWS} \times \text{DEF}$$

$$= [(\partial u / \partial z)^2 + (\partial v / \partial z)^2]^{1/2} \times [(\partial u / \partial x - \partial v / \partial y)^2 + (\partial v / \partial x - \partial u / \partial y)^2]^{1/2}, \quad (2)$$

$$\text{DVT} = C |(\partial u / \partial x + \partial v / \partial y)_{t_2} - (\partial u / \partial x + \partial v / \partial y)_{t_1}|, \quad (3)$$

where u and v are the zonal and meridional components of the horizontal wind, respectively. Here, VWS is calculated by the finite difference method using the geopotential height z at each pressure level, and the result is similar to that using the chain rule (e.g., $du/dz = du/dp \times dp/dz$) based on the hydrostatic assumption, for both the fine resolution model (e.g., ECMWF) and the coarse resolution model (e.g., CPTEC). DVT requires the model fields from two different valid times represented as subscripts of t_1 and t_2 in Eq. (3). In this study, the model fields from 24-h (t_1) and 30-h (t_2) forecasts are used for determining DVT valid at $T + 30$ h. The term C is an empirical constant that makes the magnitudes of DVT equivalent to that of Ellrod1 in an environmental condition favorable for CAT generation associated with IGWs (Ellrod and Knox 2010). To determine the best value of C , we constructed probability density functions (PDFs) for Ellrod1 and DVT at 250 hPa in the northern midlatitude regions (20°–60°N latitude) where Ellrod1 and DVT are dominant using the seven NWP model outputs for the evaluation period, which is shown in Fig. 1. The peak values of the PDF bins for Ellrod1 appear at around $\sim 10^{-7} \text{ s}^{-2}$, whereas those for DVT appear at around $\sim 10^{-5} \text{ s}^{-2}$, implying that the magnitudes of DVT are generally 100 times larger than the magnitudes of Ellrod1 in all models. Therefore, in this study, the constant C is assigned as 0.01, which differs from Ellrod and Knox (2010) where C is assigned as 0.1. Note that C is sensitive to the model resolution.

As in many previous studies (e.g., Sharman and Pearson 2017; Kim et al. 2018), DVT can be estimated from the instantaneous divergence tendency equation, which might be beneficial for operation forecasts by saving of computation time and data storage space. When we additionally calculated EKI using the instantaneous divergence tendency equation, the performance skills of the multimodel-based ensemble CAT forecasts are slightly lower (not shown) than those of the current result.

FIG. 1. PDFs for 30-h (a) Ellrod1 and (b) DVT forecasts at 250 hPa in the 20°–60°N latitude band calculated by using seven TIGGE NWP model outputs for a 6-month period (October 2016–March 2017).

Figure 2 shows an example of the single-model-based EKI forecasts at 250 hPa valid at 1800 UTC 31 December 2016, derived from the seven TIGGE model fields. All forecasts predict strong CAT near upper cyclonic and anticyclonic flows in which strong deformation, vertical wind shear, and divergent flow exist. In particular, stronger CAT areas appear over the North and South Pacific Oceans, North Atlantic Ocean, southern Greenland, and Alaska. Although the patterns of the EKI forecasts look similar to each other, there are some minor discrepancies in local areas. This is likely because of differences in the wind forecasts depending on the model resolution, dynamic core, data assimilation system, and physical packages. In particular, the EKI forecasts from the ECMWF, the Korea Meteorological Administration (KMA), and the Met Office (UKMO) show stronger CAT areas near Alaska than other models (Figs. 2c,e,g). On the other hand, the EKI forecast from the Brazil Centro de Previsão de Tempo Estudos Climático (CPTEC), which has relatively coarse grid spacing in its underlying model (Table 1), shows very weak CAT areas than the others (Fig. 2b).

c. Multimodel-based ensemble CAT forecasts

In this study, we suggest two types of the multimodel-based ensemble CAT forecasts using the seven NWP-model-derived EKIs. First is an ensemble mean product of all EKI forecasts [multimodel-based ensemble mean (MMEM)], as follows:

$$\text{MMEM}(x, y, z) = \left[\sum_{i=1}^N \text{EKI}_i(x, y, z) \right] / N, \quad (4)$$

where, N is the number of models in the MMEM, which is seven in the present case. Second is a multimodel-based probabilistic (MMP) CAT forecast. The MMP is generated by calculating a relative percentage of EKIs

that exceed a certain threshold of EKI for moderate-or-greater (MOG)-level turbulence among EKIs derived from seven TIGGE model outputs at a given grid as follows (Gill and Buchanan 2014; Kim et al. 2018; Storer et al. 2019):

$$\text{MMP}(x, y, z) = \left[\sum_{m=1}^{N_t} N_m(x, y, z) \right] / N_t \times 100, \quad (5)$$

where N_t is the number of models in the MMP, which is seven in the present study, and $N_m(x, y, z)$ is a binary outcome for detecting turbulence. If EKI calculated using the m th NWP model exceeds a certain EKI threshold at a given grid point, $N_m(x, y, z)$ is set to be 1, otherwise 0. The EKI threshold is determined by considering the climatological frequency of MOG turbulence encounters obtained from the long period of turbulence observation data (Sharman et al. 2006; Wolff and Sharman 2008). Therefore, this can be translated as the likelihood of MOG turbulence encounters based on the EKI climatology. Sharman et al. (2014) found that the climatological frequency of the upper-level MOG turbulence observed from in situ eddy dissipation rate (EDR) measurements during 2009–13 was about 0.1%, while Wolff and Sharman (2008) found that the climatological frequency of the upper-level MOG turbulence observed from pilot reports (PIREPs) during 1994–2005 was less than 2%. In this study, upper-level ($>20,000$ ft) MOG turbulence ($\text{EDR} > 0.22 \text{ m}^{2/3} \text{ s}^{-1}$) observed from in situ EDR during the research period was about 0.1%, which is much less than that from PIREPs.

Figures 3a–g show global distributions of the 6-month mean for the EKI forecasts at 250 hPa valid at 1800 UTC calculated using the seven NWP models. The global distributions of the 6-month mean for the Ellrod1 and DVT forecasts are given, respectively, in Figs. S1 and S2 in the online supplemental material. In general, the

FIG. 2. The 30-h EKI forecasts at 250 hPa derived from seven TIGGE NWP model fields valid at 1800 UTC 31 Dec 2016.

EKI values are stronger in the midlatitude region than in other areas in both Northern (NH) and Southern (SH) Hemispheres, largely due to its association with upper-level jet/frontal systems. The EKI magnitudes are stronger in the NH, due to the NH winter period.

Another noticeable feature in Figs. 3a–g is that the EKI magnitudes (intensities) are quite different from model to model, although the overall features of strong EKI areas are similar. For example, in the midlatitudes, the EKI values are the strongest in the ECMWF (Fig. 3c)

FIG. 3. (a)–(g) Global distributions and (h) PDFs of 30-h EK1 forecasts at 250 hPa derived from seven TIGGE NWP model outputs for a 6-month period (October 2016–March 2017). The PDFs are constructed using values of EK1 in the 20°–60°N latitude band at 250 hPa.

while it is the weakest in the CPTEC (Fig. 3b), which is highly correlated with the horizontal resolution of the underlying model (Table 1). This suggests that the threshold of EK1 representing MOG turbulence for each NWP model should be different.

Considering that the climatological frequency of MOG turbulence in the upper-level region from the observation data was less than 2% (Wolff and Sharman 2008; Sharman et al. 2014), the 98th percentile values of the seven EK1s are used as the ideal MOG threshold for

TABLE 2. The 98th percentile values of PDFs of EKI calculated from seven TIGGE models for a 6-month period (October 2016–March 2017) in the 20°–60°N latitude band.

Model	Threshold ($\times 10^{-6}$)	Model	Threshold ($\times 10^{-6}$)
CMA	1.298 22	CPTEC	0.616 25
ECMWF	1.610 18	JMA	1.029 09
KMA	1.525 50	NCEP	1.227 48
UKMO	1.552 94		

CAT in this study. To determine the 98th percentile value for each model, PDFs of EKIs at 250 hPa in the 20°–60°N latitude band were constructed (Fig. 3h). EKI values for all seven models range from 10^{-8} to 10^{-6} s^{-2} , but peaks are somewhat different. For example, the PDF for the ECMWF model (green) trends toward higher values of EKI (right side, Fig. 3h), while that for the CPTEC model (gray) trends toward lower values of EKI (left side, Fig. 3h), which is expected from the differences in their resolutions (Table 1 and Figs. 3b,c). The 98th percentile values of EKI are described in Table 2.

To examine the sensitivity of the EKI threshold to the MMP forecasts, different percentile values of EKI for MOG turbulence are also considered, based on the previous observational studies of in situ EDR and PIREP data: Sharman et al. (2006) of less than 1% in USA, and Kim and Chun (2011) and Lee and Chun (2015) of 3%–4% over East Asia and South Korea, respectively. Considering these previous studies, the 99th and 95th percentile values of EKI are additionally tested.

Figure 4 shows an example of the MMEM and MMP forecasts at 250 hPa valid at the same time as shown in Fig. 2. Here, the MMP is based on the 98th percentile EKI value (Table 2). In Fig. 4a, strong CAT areas in the MMEM forecast appear along the sharply curved upper-level trough and ridge areas in the midlatitudes, which is similar to the single-model-based EKI forecasts (Fig. 2). In Fig. 4b, the MMP forecast also predicts high CAT potentials near northwestern Canada, southern Greenland, across the North Atlantic Ocean, and in the Pacific Ocean, which coincide with the representative regions of strong CAT in the MMEM forecast (Fig. 4a). It is also noticeable that the MMP forecast excludes large areas of lower EKI value in the MMEM forecast, which could help reduce false alarms. Additionally, if the lower (higher) EKI threshold like the 95th (99th) percentile is applied in the MMP, larger (smaller) CAT potential areas appear than that using the 98th percentile EKI threshold in Fig. 4b (not shown).

Figure 5 shows the MMEM and MMP forecasts at 250 hPa over the North Atlantic region valid at the same

FIG. 4. As in Fig. 2, but for the (a) MMEM and (b) MMP forecasts based on EKI derived from seven TIGGE NWP model outputs, superimposed on the geopotential height (black contours; 150-gpm interval) obtained from the ECMWF 30-h forecast.

time shown in Fig. 4, with the observed turbulence. DEF, VWS, and DVT in the same area are also shown in Fig. 5. The sharply curved upper-level trough–ridge couplet is located over the North Atlantic region. A strong westerly jet with a maximum wind speed of over 90 m s^{-1} along the ridge near southern Greenland produces strong DEF (Fig. 5c) and VWS (Fig. 5d) in the vicinity of jet entrance (45° – 70°N , 55° – 60°W) and exit (55° – 65°N , 20°W) regions especially in its cyclonic shear side, which is consistent with very high values of the EKI forecasts in Figs. 5a and b. In addition, strong DVT (Fig. 5e) is occurred along this anticyclonic curvature of the flow near southern Greenland (65° – 72°N , 30° – 50°W) and in the backside of the cutoff low in the middle of the North Atlantic Ocean (40° – 50°N , 35°W), which makes enhancing areas of EKI. Both the MMEM and MMP forecasts capture generally well the observed MOG turbulence in these regions (shown as blue asterisks). In a comparison of Fig. 5a with Fig. 5b, the MMP forecast (Fig. 5b) excludes the low-risk region and isolates well the high-risk region of CAT revealed in the MMEM forecast (Fig. 5a). Because of this, the MMP forecast seems to miss a few MOG turbulence events in the regions where the

FIG. 5. (a),(b) As in Fig. 4, but with wind barbs for the focal region zoomed in near Greenland over the North Atlantic. For analysis, the 30-h ensemble-mean forecasts of (c) DEF, (d) VWS, and (e) DVT derived from seven TIGGE NWP model outputs are also represented in the figure. NIL and MOG turbulence events reported within 2 h centered at the valid time at $z = 29\,000\text{--}35\,000$ ft from in situ EDR and PIREPs are depicted by black dots and blue asterisks, respectively.

CAT areas predicted from the MMEM forecast is relatively weak. However, false alarms (shown as black dots) are much reduced compared to the MMEM forecast. These results suggest that the probabilistic forecast could benefit aviation weather users to take advantage of better forecasts for both MOG and null (NIL) events. Additionally, when the lower (higher) EKI threshold is used, the MMP forecast predicts correctly more (less) MOG events and less (more) NIL events (not shown). The prediction skill of the MMP forecast for NIL and MOG events depending on the EKI threshold will be discussed further details in Fig. 9.

In this study, we additionally constructed and evaluated the MMP forecast using the weighting strategy as a part of the method to improve the performance skill of the MMP forecast. The weighing strategy is a method of applying weights based on the performance skill of each deterministic EKI forecast, which is used in previous studies when multiple diagnostics are combined to improve the forecast performance skill (Sharman et al. 2006; Kim et al. 2011). The performance skill of the deterministic EKI forecast is represented by a value of the area under the relative operating characteristic (ROC) curve (AUC), which will be shown in Fig. 7.

The details of the ROC curve and AUC will be described in section 3. The MMP forecast with weights (MMPW) can be obtained by the following equation:

$$\text{MMPW}(x, y, z) = \left[\sum_{m=1}^{N_t} W_m N_m(x, y, z) \right] / N_t \times 100, \quad (6)$$

where $N_m(x, y, z)$ and N_t are the same variables as in Eq. (5), and W_m is a weight for the m th TIGGE model based on a value of its AUC value (AUC_m), given as $W_m = \text{AUC}_m / \left[\sum_{m=1}^{N_t} (\text{AUC}_m) / N_t \right]$. Note that in the original MMP equation, W_m is set to be 1. The prediction skill of the MMPW forecast for NIL and MOG events will be also discussed in Fig. 9.

d. Turbulence observations

To conduct statistical evaluations of two multimodel-based ensemble CAT forecasts, the two resources of turbulence observations are considered. The first is PIREPs, which are routine or event-based manual weather reports from pilots with detailed information about observation time, longitude, latitude, flight altitude, and intensity of atmospheric hazards including

FIG. 6. Horizontal distributions of (a),(c) in situ EDR (NIL: $\text{EDR} < 0.01 \text{ m}^{2/3} \text{ s}^{-1}$; MOG: $\text{EDR} > 0.22 \text{ m}^{2/3} \text{ s}^{-1}$) and (b),(d) PIREPs reported at $z = 29\,000\text{--}35\,000$ ft within 2 h centered at 1800 UTC for a 6-month period (October 2016–March 2017) Blue dots in (a) and (b) and red dots in (c) and (d) represent NIL and MOG turbulence events, respectively.

turbulence. Although PIREPs have some uncertainties in location and intensity of the reported turbulence, it has been used for long-term evaluations of turbulence forecasts in previous studies (e.g., Sharman et al. 2006; Wolff and Sharman 2008; Kim et al. 2011; Wandishin et al. 2011).

We also used in situ EDR observations, which are automatically recorded by some commercial aircrafts using an onboard EDR algorithm based on vertical acceleration or vertical velocity (Cornman et al. 2004; Sharman et al. 2014). In situ EDR observations are archived regularly at every 1 or 15 min and are sometimes reported for triggered events detected when the EDR value exceeds a certain threshold (Sharman et al. 2014). In the data, time, location, and 1-min peak and mean values of the EDR estimates that can be correlated to the turbulence intensity are provided (Sharman et al. 2014; Sharman and Pearson 2017; Kim et al. 2018).

In this study, NIL and MOG turbulence events observed from in situ EDR and PIREPs reported within 2 h centered at 1800 UTC at $z = 29\,000\text{--}35\,000$ ft (8.8–10.7 km) are used to evaluate two products of the multimodel-based ensemble CAT forecasts. In situ EDR and PIREPs were provided from the National Center for Atmospheric Research (NCAR) for research purposes. To categorize turbulence intensity from in situ EDR observations, thresholds of 0.01 and $0.22 \text{ m}^{2/3} \text{ s}^{-1}$ are chosen to represent the NIL and MOG intensity of turbulence, respectively, following Kim et al. (2018).

Figure 6 shows horizontal distributions of NIL and MOG turbulence events from in situ EDR (Figs. 6a,c) and PIREPs (Figs. 6b,d) used for the evaluation. As shown in Fig. 6, NIL and MOG turbulence events from in situ EDR and PIREPs are observed along the major international flight routes from/to United States to/from Europe, Africa, South America, and Asia. Most of these observations are reported in the United States, while few are observed in the SH, Eurasia, and Africa. Most commercial in situ EDR observations are proprietary and may be difficult to obtain for research purposes. The numbers and relative percentages of NIL and MOG turbulence events obtained from in situ EDR and PIREPs shown in Fig. 6 are summarized in Table 3. Here, the relative percentage is the number of NIL (or MOG) events divided by total number (NIL + MOG) of each observation data used in the evaluation. Because automated in situ EDR observations are frequently reported in a regular basis (every 1 or 15 min), they have a large number of NIL events (about 21 times higher than that from PIREPs), while PIREPs have relatively small number of NIL events (Fig. 6a versus Fig. 6b and Table 3). Conversely, the number of MOG turbulence events from PIREPs is about 10 times larger than that from in situ EDR (Fig. 6c versus Fig. 6d and Table 3). Even though PIREPs include the uncertainty of the turbulence reports, adding them can help more robust evaluations of the multimodel-based ensemble CAT forecasts by using more MOG turbulence events.

TABLE 3. Numbers corresponding to NIL and MOG turbulence reported at $z = 29\,000\text{--}35\,000$ ft within 2 h centered at 1800 UTC for a 6-month period (October 2016–March 2017), which are sorted for statistical evaluations of the MMEM and MMP forecasts.

	In situ EDR No. (percentage)	PIREP No. (percentage)
Null (NIL)	119 862 (99.7%)	2307 (41.4%)
Moderate-or-greater (MOG)	314 (0.3%)	3266 (58.6%)
Total	120 176 (100%)	5573 (100%)

3. Evaluation of the multimodel-based ensemble CAT forecasts

In this study, the MMEM and MMP CAT forecast products are suggested by computing EKI from the seven TIGGE model outputs through the methodologies described in section 2. For the objective evaluations, the 30-h forecasts of these two products valid at 1800 UTC at 250 hPa and available turbulence observation data for a 6-month period (October 2016–March 2017) are used.

Additionally, we examined the sensitivity of turbulence diagnostics to the MMEM and MMP forecasts considering two additional CAT diagnostics, the inverse of the Richardson number ($1/Ri$) and the NCSU index 1 (NCSU1) (Kaplan et al. 2006), which can predict regions of turbulence different from those of EKI.

a. Evaluation methods

In this study, ROC curves based on probability of detection (POD) statistics (Mason 1982) are used to evaluate the performance skill of the MMEM forecast, which have been used as an objective, statistical evaluation tools for turbulence forecasts in the previous studies (e.g., Sharman et al. 2006; Kim et al. 2011; Gill 2014; Sharman and Pearson 2017; Kim et al. 2018). According to the POD statistics, the probabilities of the turbulence forecast which detects MOG events as “yes” (POD_y) and the probabilities of the turbulence forecast which detects NIL events as “no” (POD_n) are derived by varying thresholds of the turbulence forecast. Then, the ROC curves are constructed from the multiple pairs of POD_y and POD_n (probability of false alarm detection = $1 - \text{POD}_n$). An AUC value of the ROC curve is finally used as a performance skill score of the turbulence forecast.

For the evaluation of the MMP forecast, a reliability diagram (Gill and Buchanan 2014; Sharman and Lane 2016; Kim et al. 2018; Storer et al. 2019) is constructed, which is drawn by calculating the frequency of the observed turbulence encounters within a given forecast probability. If the reliability line is located below (above) the perfect reliability line depicted as the diagonal line, then the probabilistic forecast tends to be

overforecasting (underforecasting) (Gill and Buchanan 2014; Kim et al. 2018; Storer et al. 2019).

b. Statistical evaluation of MMEM forecasts

Figure 7 shows the ROC curves for the deterministic EKI forecasts derived from single models and MMEM. In this evaluation, in situ EDR and PIREPs are both used. In Fig. 7, among the seven single-model-based EKI forecasts, the UKMO and CPTEC model-based EKI forecasts have the highest (AUC = 0.765) and the lowest (AUC = 0.722) performance skills, respectively. The ECMWF and KMA model-based EKI forecasts also have good performance skills comparable to the UKMO. Differences in the performance skills between the single-model-based EKI forecasts could be related to the differences in native model resolution (Table 1) and model physics. In particular, the wind fields near cyclonic and anticyclonic flows, which are important to generate gravity waves through the flow imbalance, are dependent on the differences in model physics as well as model resolution.

The AUC value of the MMEM forecast (denoted by EnM7 in Fig. 7) is 0.805 and is higher than the AUC value of the best single-model-based EKI forecast (UKMO). This demonstrates that combining the EKI diagnostics calculated based on multiple NWP models also has the improved forecast skill compared to the single-model-based EKI diagnostic. Storer et al. (2019) showed that the AUC for the mean of the probabilistic forecasts based on two different ensemble model data is better than that for either of two single-ensemble-based probabilistic forecasts. This is consistent with the result of this study.

To consider the irregular nature of location and frequency of turbulence observations and show the statistical significance in ROC-based performance skills, a sensitivity test using observation subsets is additionally performed (Sharman et al. 2006; Kim et al. 2011). First, a subset of turbulence observations is constructed by randomly selecting half of the available turbulence observations (PIREPs and in situ EDR). Then, for a subset, a ROC curve is derived. For the sensitivity test, this process is repeated 1000 times in this study. The dashed and dotted curves in Fig. 7 show the two ROC curves with maximum and minimum AUC values among the 1000 ROC curves. The results fall in the range from -3.1% to $+2.1\%$ around the individual AUC values against the total observations, which represents a statistically robust and stable performance skill for all evaluation tests. This is quite similar to the results in Kim et al. (2011, 2018) and Sharman et al. (2006). The results shown in Fig. 7 are generally consistent with those calculated using $1/Ri$ and NCSU1 (Fig. S3).

FIG. 7. Statistical performances (thick solid line), which are indicated by PODY–POFD lines with the AUC skill scores (lower right), for the 30-h EKI forecasts derived from each of seven TIGGE NWP models and the 30-h MEM (EnM7) forecast. The evaluation period is six months (October 2016–March 2017). For the evaluation, NIL and MOG turbulence observations from in situ EDR and PIREPs are used. The maximum and minimum boundaries among the 1000 ROC curves derived using the subsets constructed by randomly selecting half of in situ EDR and PIREPs are represented by dashed and dotted lines, respectively. The diagonal line is the no-skill line.

To examine the sensitivity of the observation data type to the performance skill, the ROC curves for the deterministic EKI forecasts from single models and MEM are calculated against each of in situ EDR and PIREPs, and the results are shown in Fig. 8. It is found that the performance skill of the MEM forecast (EnM7) is superior to any other single-model-based EKI forecasts against not only in situ EDR but also PIREPs, which is identical to the result shown in Fig. 7.

Also, the AUC values are higher when the EKI forecast is verified against in situ EDR, because it is an objective source of observational data.

c. Statistical evaluation of MMP forecasts

Figure 9 shows the statistical reliability diagrams (thick solid lines) for three MMP forecasts derived using the thresholds of the 95th, 98th, and 99th percentile values of EKIs. The bar chart represents the numbers of

FIG. 8. As in Fig. 7, but for evaluating the deterministic CAT forecasts against each of in situ EDR and PIREPs.

NIL (gray bars) and MOG (white tips on top of gray bars) turbulence events from in situ EDR and PIREPs used to compute the observed frequency in each forecast probability of three MMP forecasts. The black dotted line denotes the sample climatology calculated by a ratio value of the number of the turbulence events to the number of forecasts. The reliability line for the MMP farther away from the sample climatology line has a better resolution. Figure 9 shows that the MMP forecasts with EKI thresholds of the high-percentile values are more reliable and have a better resolution. This is because using the high-percentile value as the threshold makes the smaller areas of probabilistic forecast (i.e.,

less overforecasting). Similar results of MMP are also shown for the 1/Ri and NCSU1 (see Fig. S4). At the low forecast probabilities of less than 20%, the probabilistic forecasts show better reliability, while those show an overforecasting for MOG turbulence events at the higher forecast probabilities. Especially, at forecast probabilities from 30% to 90%, the slope of reliability is much lower than the perfect reliability line, which can be attributed for the following possible reasons.

- 1) Mismatch of the scales between the observed turbulence and EKI forecast: The size of turbulent eddies experienced by aircraft can range from 10 to 1000 m (Sharman et al. 2014), which is still very smaller than

FIG. 9. Reliability diagrams (thick solid lines) for three 30-h MMP forecasts valid at 1800 UTC for a 6-month period (October 2016–March 2017). The three MMP forecasts are derived by applying the threshold of the 95th (black line), 98th (red line), and 99th (blue line) percentile values of seven TIGGE model-based EKIs at 250 hPa in the northern midlatitude region for the evaluation period. The 99th percentile EKI-based MMPW forecast is represented by the blue dashed line. The numbers of NIL (gray bars) and MOG (white tips on top of gray bars) turbulence events captured in each forecast probability of the 95th (dark gray), 98th (gray), and 99th (light gray) percentile EKI value-based MMP forecasts are depicted by the bar chart. For better comparison, the numbers of NIL and MOG turbulence events in the 0% probability are multiplied by 1/5. The diagonal (black thin) and the horizontal (black dotted) lines represent the perfect reliability and the sample climatology lines, respectively.

the model resolution used in this study (Table 1). This can result in overforecasting (Kim et al. 2018).

- 2) Limitation of single turbulence diagnostic: only EKI is used in the determination of the MMP, which may miss turbulence from other sources, such as convection and mountain waves. The turbulence forecasts additionally produced by other diagnostics may capture more turbulence encounters, which can affect the reliability for MMP.
- 3) Lack of spread in the MMP forecast: the global NWP models generally have small spread in short-term forecasts (<48 h), because their initial perturbations are tuned to forecast for seasonal-to-subseasonal (S2S) projections (2 weeks). In addition, because only seven deterministic NWP outputs are used to compute the MMP, the final MMP forecasts have less spread.
- 4) Lack of observed MOG events: actual turbulence information is only obtained from the aircraft reports, which suffers inherent avoidance bias because aircraft is trying to avoid known turbulence regions. This bias contributes to a small observed frequency at each probability due to fewer MOG events compare

to NIL events (bar chart in Fig. 9). For the sensitivity test to the number of observed turbulence events, the performances of the MMP forecasts are also evaluated against light-or-greater (LOG)-level turbulence events (not shown) classified by the criteria of $\text{EDR} > 0.15 \text{ m}^{2/3} \text{ s}^{-1}$ (Sharman and Pearson 2017). It is found that as more turbulence events (857) are added, the reliability and the resolution of all MMP forecasts become better at predicting LOG turbulence events than MOG turbulence events, which is consistent with Kim et al. (2018).

The MMPW forecast, which includes AUC weighting on the MMP forecast, is additionally examined for improving the performance skill of the MMP forecast. The evaluation result of the MMPW forecast using the thresholds of the 99th percentile EKI values (blue dashed line) is shown in Fig. 9. The MMPW forecast, which is still in the right side of the diagonal line (i.e., overforecasting), is very similar to the MMP forecast, although with a slightly more (less) reliable at the low (high) forecast probabilities.

4. Summary and conclusions

In this study, we evaluated two multimodel-based ensemble forecasts of CAT with a single diagnostic EKI, named MMEM and MMP, using the seven global NWP model outputs from the TIGGE database. The MMEM forecast is computed by the ensemble mean of the seven EKIs at each grid point. The MMP forecast, which represents the probability of encountering MOG turbulence, is produced by calculating the relative percentage agreement of how often EKIs exceed a predetermined MOG EKI threshold among the seven single-model-based EKI values at each grid point. To examine the performance skill of the MMP when the EKI threshold is applied differently, the 95th, 98th, and 99th percentile values of each of the seven EKI PDFs are tested.

In the case study (Figs. 4 and 5), strong CAT regions predicted from the MMEM and MMP 30-h forecasts at 250 hPa appear in the cyclonic shear side near the entrance and exit regions of the upper-level jet by strong DEF and VWS, and also in anticyclonic curvature and backside of cutoff low regions by strong DVT. The locally enhanced CAT regions in the two MMEM and MMP forecasts are consistent with the locations of the observed MOG turbulence events over the North Atlantic. The MMP forecast using the high EKI threshold isolates the strong CAT regions, which could be more beneficial for correctly detecting both MOG turbulence and NIL events.

Based on the POD method used for deriving a ROC curve and the reliability diagram, the performance skills of the TIGGE-model-based ensemble 30-h CAT forecasts are evaluated against available turbulence observations over the global region. In the evaluation, NIL and MOG turbulence events reported within 2 h centered at 1800 UTC at $z = 29\,000\text{--}35\,000$ ft from PIREPs and in situ EDR observations are considered. In the evaluation result of the deterministic forecasts using the POD method, the KMA, ECMWF, and UKMO show better performance skills among the single-model-based EKI forecasts. This result is highly correlated to the underlying model resolution, and other factors like physics packages. When considering MMEM, the performance skill of the MMEM forecast is better than any single-model-based EKI forecasts, implying that the ensemble mean of EKIs using multiple NWP models can improve the performance skill of predicting turbulence encounters.

In the evaluation result of the MMP forecasts using three different percentile-based EKI thresholds using the reliability diagram, using the high-percentile values of EKI thresholds has better reliability, although all MMP forecasts show an overforecasting. Additionally, the MMP forecast with the weight (MMPW), which accounts for each NWP model's performance (AUC), is conducted and is evaluated by the observation data. The reliability of MMPW is nearly the same as for that of MMP. Overforecasting of MMP in the reliability is likely due to several potential reasons. First, the scales between the observed turbulence events and the NWP model-derived CAT forecasts are mismatched. Second, because the MMP forecasts are derived using only a single CAT diagnostic EKI, CAT potential regions generated by various sources are not represented sufficiently in MMP. Third, because only seven short-term (30-h) forecasts from different NWP models are considered in the present study, MMP has a limit to provide better spread. Fourth, turbulence events in observation data are significantly less than in real atmosphere due to the inherent bias from commercial aircraft that is trying to avoid known turbulence regions. The results for EKI are found to be similar to those for two additional turbulence diagnostics, $1/R_i$ and NCSU1.

This study suggests that the use of a multimodel ensemble in the development of turbulence forecasts will be eventually beneficial for aviation users in the world. To further improve the performance of the multimodel-based turbulence forecasts, it would be necessary to consider multiple diagnostics related to other sources (Kim et al. 2018). Because considering several diagnostics together can provide more areas of turbulence, this way will make a more reliable forecast spread with better accuracy, which can benefit several operational providers.

Acknowledgments. This work is based on TIGGE data. We thank ECMWF and CMA for providing valuable ensemble forecast data and supporting user scientific research via data archive portals. PIREP and in situ EDR observations were provided from NCAR for research purposes. This work was funded by the Korea Meteorological Administration Research and Development Program under Grant KMI2018-07810. Jung-Hoon Kim was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2019R111A2A01060035).

REFERENCES

- Bougeault, P., and Coauthors, 2010: The THORPEX Interactive Grand Global Ensemble. *Bull. Amer. Meteor. Soc.*, **91**, 1059–1072, <https://doi.org/10.1175/2010BAMS2853.1>.
- Cornman, L. B., G. Meymaris, and M. Limber, 2004: An update on the FAA Aviation Weather Research Program's in situ turbulence measurement and reporting system. *11th Conf. on Aviation, Range, and Aerospace Meteorology*, Hyannis, MA, Amer. Meteor. Soc., P4.3, <https://ams.confex.com/ams/pdfpapers/81622.pdf>.
- Dutton, J. A., and H. A. Panofsky, 1970: Clear air turbulence: A mystery may be unfolding. *Science*, **167**, 937–944, <https://doi.org/10.1126/science.167.3920.937>.
- Ellrod, G. P., and D. I. Knapp, 1992: An objective clear-air turbulence forecasting technique: Verification and operational use. *Wea. Forecasting*, **7**, 150–165, [https://doi.org/10.1175/1520-0434\(1992\)007<0150:AOCATF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1992)007<0150:AOCATF>2.0.CO;2).
- , and J. A. Knox, 2010: Improvements to an operational clear-air turbulence diagnostic index by addition of a divergence trend term. *Wea. Forecasting*, **25**, 789–798, <https://doi.org/10.1175/2009WAF2222290.1>.
- Gill, P. G., 2014: Objective verification of World Area Forecast Centre clear air turbulence forecasts. *Meteor. Appl.*, **21**, 3–11, <https://doi.org/10.1002/met.1288>.
- , and P. Buchanan, 2014: An ensemble based turbulence forecasting system. *Meteor. Appl.*, **21**, 12–19, <https://doi.org/10.1002/met.1373>.
- Guest, F. M., M. J. Reeder, C. J. Marks, and D. J. Karoly, 2000: Inertia-gravity waves observed in the lower stratosphere over Macquarie Island. *J. Atmos. Sci.*, **57**, 737–752, [https://doi.org/10.1175/1520-0469\(2000\)057<0737:IGWOIT>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<0737:IGWOIT>2.0.CO;2).
- Kaplan, M. L., A. W. Huffman, K. M. Lux, J. J. Charney, A. J. Riordan, and Y.-L. Lin, 2005a: Characterizing the severe turbulence environments associated with commercial aviation accidents. Part 1: A 44-case study synoptic observational analyses. *Meteor. Atmos. Phys.*, **88**, 129–152, <https://doi.org/10.1007/s00703-004-0080-0>.
- , —, J. D. Cetola, J. J. Charney, A. J. Riordan, Y.-L. Lin, and K. T. Waight III, 2005b: Characterizing the severe turbulence environments associated with commercial aviation accidents. Part 2: Hydrostatic mesoscale numerical simulations of supergradient wind flow and streamwise ageostrophic frontogenesis. *Meteor. Atmos. Phys.*, **88**, 153–173, <https://doi.org/10.1007/s00703-004-0079-6>.
- , and Coauthors, 2006: Characterizing the severe turbulence environments associated with commercial aviation accidents. A real-time turbulence model (RTTM) designed for

- the operational prediction of hazardous aviation turbulence environments. *Meteor. Atmos. Phys.*, **94**, 235–270, <https://doi.org/10.1007/s00703-005-0181-4>.
- 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, <https://doi.org/10.1175/2010JAMC2449.1>.
- , and —, 2011: Statistics and possible sources of aviation turbulence over South Korea. *J. Appl. Meteor. Climatol.*, **50**, 311–324, <https://doi.org/10.1175/2010JAMC2492.1>.
- , and —, 2012: Development of the Korean aviation Turbulence Guidance (KTG) system using the operational Unified Model (UM) of the Korea Meteorological Administration (KMA) and Pilot Reports (PIREPs). *J. Korean Soc. Aviat. Aeronaut.*, **20**, 76–83, <https://doi.org/10.12985/ksaa.2012.20.4.076>.
- , —, 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, <https://doi.org/10.1175/JAMC-D-10-05017.1>; Corrigendum, **50**, 2193, <https://doi.org/10.1175/JAMC-D-11-0188.1>.
- , W. Chan, B. Sridhar, R. Sharman, P. Williams, and M. Strahan, 2016: Impact of the North Atlantic oscillation and transatlantic flight routes and clear-air turbulence. *J. Appl. Meteor. Climatol.*, **55**, 763–771, <https://doi.org/10.1175/JAMC-D-15-0261.1>.
- , R. Sharman, M. Strahan, J. W. Scheck, C. Bartholomew, J. C. H. Cheung, P. Buchanan, and N. Gait, 2018: Improvements in non-convective aviation turbulence prediction for the World Area Forecast System (WAFS). *Bull. Amer. Meteor. Soc.*, **99**, 2295–2311, <https://doi.org/10.1175/BAMS-D-17-0117.1>.
- Knox, J. A., 1997: Possible mechanisms of clear-air turbulence in strongly anticyclonic flows. *Mon. Wea. Rev.*, **125**, 1251–1259, [https://doi.org/10.1175/1520-0493\(1997\)125<1251:PMOCAT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1251:PMOCAT>2.0.CO;2).
- , 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, <https://doi.org/10.1175/2008JAS2477.1>.
- Lee, D.-B., and H.-Y. Chun, 2015: A statistical analysis of aviation turbulence observed in pilot report (PIREP) over East Asia including South Korea. *Atmosphere*, **25**, 129–140, <https://doi.org/10.14191/ATMOS.2015.25.1.129>.
- Lee, D. R., R. S. Stull, and W. S. Irvine, 1984: Clear air turbulence forecasting techniques. Air Weather Service Tech. Note AFGWC/TN-79/001 (REV), 86 pp.
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Plougonven, R., H. Teitelbaum, and V. Zeitlin, 2003: Inertia gravity wave generation by the tropospheric midlatitude jet as given by the fronts and Atlantic storm-track experiment radio soundings. *J. Geophys. Res.*, **108**, 4686, <https://doi.org/10.1029/2003JD003535>.
- Sharman, R. D., and T. Lane, Eds., 2016: *Aviation Turbulence: Processes, Detection, Prediction*. Springer, 523 pp.
- , and J. Pearson, 2017: Prediction of energy dissipation rates for aviation turbulence. Part I: Forecasting nonconvective turbulence. *J. Appl. Meteor. Climatol.*, **56**, 317–337, <https://doi.org/10.1175/JAMC-D-16-0205.1>.
- , C. Tebaldi, G. Wiener, and J. Wolff, 2006: An integrated approach to mid- and upper-level turbulence forecasting. *Wea. Forecasting*, **21**, 268–287, <https://doi.org/10.1175/WAF924.1>.
- , S. B. Trier, T. P. Lane, and J. D. Doyle, 2012: Sources and dynamics of turbulence in the upper troposphere and lower stratosphere: A review. *Geophys. Res. Lett.*, **39**, L12803, <https://doi.org/10.1029/2012GL051996>.
- , L. B. Cornman, G. Meymaris, and J. Pearson, 2014: Description and derived climatologies of automated in situ eddy-dissipation-rate reports of atmospheric turbulence. *J. Appl. Meteor. Climatol.*, **53**, 1416–1432, <https://doi.org/10.1175/JAMC-D-13-0329.1>.
- Storer, L. N., P. G. Gill, and P. D. Williams, 2019: Multi-model ensemble predictions of aviation turbulence. *Meteor. Appl.*, **26**, 416–428, <https://doi.org/10.1002/MET.1772>.
- Swinbank, R., and Coauthors, 2016: The TIGGE project and its achievements. *Bull. Amer. Meteor. Soc.*, **97**, 49–67, <https://doi.org/10.1175/BAMS-D-13-00191.1>.
- Uccellini, L. W., and S. E. Koch, 1987: The synoptic setting and possible energy sources for mesoscale wave disturbances. *Mon. Wea. Rev.*, **115**, 721–729, [https://doi.org/10.1175/1520-0493\(1987\)115<0721:TSSAPE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<0721:TSSAPE>2.0.CO;2).
- Wandishin, M. S., B. P. Pettegrew, M. A. Petty, and J. L. Mahoney, 2011: Quality assessment report: Graphical turbulence guidance, version 2.5. U.S. National Oceanic and Atmospheric Administration, Earth System Research Laboratory, Global Systems Division, 52 pp., <https://repository.library.noaa.gov/view/noaa/11117>.
- Williams, P. D., 2016: Transatlantic flight times and climate change. *Environ. Res. Lett.*, **11**, 024008, <https://doi.org/10.1088/1748-9326/11/2/024008>.
- , and M. M. Joshi, 2013: Intensification of winter transatlantic aviation turbulence in response to climate change. *Nat. Climate Change*, **3**, 644–648, <https://doi.org/10.1038/nclimate1866>.
- WMO, 2005: THORPEX international research implementation plan. WMO/TD-1258, WWRP/THORPEX 4, 104 pp., www.wmo.int/pages/prog/arep/wwrp/new/documents/CD_ROM_implementation_plan_v1.pdf.
- Wolff, J. K., and R. D. Sharman, 2008: Climatology of upper-level turbulence over the contiguous United States. *J. Appl. Meteor. Climatol.*, **47**, 2198–2214, <https://doi.org/10.1175/2008JAMC1799.1>.