

Prediction of Energy Dissipation Rates for Aviation Turbulence. Part I: Forecasting Nonconvective Turbulence

R. D. SHARMAN AND J. M. PEARSON

National Center for Atmospheric Research,^a Boulder, Colorado

(Manuscript received 8 June 2016, in final form 20 September 2016)

ABSTRACT

Current automated aviation turbulence forecast algorithms diagnose turbulence from numerical weather prediction (NWP) model output by identifying large values in computed horizontal or vertical spatial gradients of various atmospheric state variables (velocity; temperature) and thresholding these gradients empirically to indicate expected areas of “light,” “moderate,” and “severe” levels of aviation turbulence. This approach is obviously aircraft dependent and cannot accommodate the many different aircraft types that may be in the airspace. Therefore, it is proposed to provide forecasts of an atmospheric turbulence metric: the energy dissipation rate to the one-third power (EDR). A strategy is developed to statistically map automated turbulence forecast diagnostics or groups of diagnostics to EDR. The method assumes a lognormal distribution of EDR and uses climatological peak EDR data from in situ equipped aircraft in conjunction with the distribution of computed diagnostic values. These remapped values can then be combined to provide an ensemble mean EDR that is the final forecast. New mountain-wave-turbulence algorithms are presented, and the lognormal mapping is applied to them as well. The EDR forecasts are compared with aircraft in situ EDR observations and verbal pilot reports (converted to EDR) to obtain statistical performance metrics of the individual diagnostics and the ensemble mean. It is shown by one common performance metric, the area under the relative operating characteristics curve, that the ensemble mean provides better performance than forecasts from individual model diagnostics at all altitudes (low, mid-, and upper levels) and for two input NWP models.

1. Introduction

Turbulence is a well-known safety and efficiency issue for commercial and general aviation (e.g., Sharman and Lane 2016). Unexpected encounters with turbulence are responsible for numerous injuries each year, with occasional fatalities, and can be a cause of aircraft fatigue and damage. For these reasons, pilots, dispatchers, and air traffic controllers attempt to avoid known areas of elevated turbulence whenever possible. When avoidance is not feasible, for instance because of consequent disruptions to air traffic, sufficient warning may mitigate the effects on passengers and crew by ensuring seat belts are fastened. Current avoidance procedures rely on turbulence forecasts to strategically assign regions of expected elevated turbulence, or on nowcasts (which are

usually based on real-time reports of turbulence) to tactically avoid turbulence, especially in regions in or near deep convection.

Current operational turbulence forecasting methods use operational numerical weather prediction (NWP) model output to capture the large-scale features that may be conducive to aviation turbulence, and are augmented by human input and real-time observations to calibrate the forecasts to provide indications of the expected “light,” “moderate,” or “severe” turbulence levels. These categorical descriptions are obviously aircraft dependent, yet given that these forecasts must be provided to all users of the airspace, with many different aircraft types and configurations, it becomes necessary to provide information about expected turbulence intensities that are independent of aircraft type. Users can then judge what this atmospheric turbulence metric means to them, as outlined in Sharman et al. (2014).

Toward establishing an atmospheric turbulence-based forecasting system, a particularly useful metric of turbulence intensity is the rate at which turbulent kinetic energy is transformed into heat; that is, the energy

^aThe National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author e-mail: Dr. Robert D. Sharman, sharman@ucar.edu

dissipation rate (ε) or $\text{EDR} = \varepsilon^{1/3}$ (Sharman et al. 2014). EDR may be related to aircraft-specific loads (MacCready 1964; Cornman et al. 1995; Sharman et al. 2014; Cornman 2016), allowing a user who is interested in a particular aircraft to calibrate EDR in terms of their specific aircraft response. Further motivating the use of EDR is that it is the International Civil Aviation Organization (ICAO 2001) standard, and conveniently provides values in the range from near 0 (typically smooth) to 1 (typically “extreme” for most aircraft) in units of meters to the two-thirds power per second. Direct estimates of EDR are now available in situ from some airlines (Sharman et al. 2014) as well as from spectral width estimates from onboard and ground-based radar (e.g., Williams and Meymaris 2016). Thus, forecasts or nowcasts of EDR can be calibrated and verified by comparison with these in situ and remotely sensed EDR observations.

In this two-part paper, we describe a method for developing automated (strategic) forecasts and (tactical) nowcasts of EDR. In this first part, we describe the forecast method and provide statistical performance results that are based on comparisons with observations. In this forecasting technique most known sources of turbulence are implicitly included [e.g., clear-air turbulence (CAT), mountain-wave turbulence (MWT), low-level turbulence (LLT), and convectively induced turbulence (CIT) for larger convective systems], using automated diagnostics of turbulence from NWP model output with an ensemble mean of the diagnostics providing the final forecast. It is an extension of the Graphical Turbulence Guidance (GTG) product described in Sharman et al. (2006) in three ways: 1) it provides forecasts at all flight altitudes, from the surface to flight level (FL) 500; 2) it provides explicit forecasts of MWT; and 3) the output is EDR instead of a likelihood metric. The frequency and lead times of the forecasts depend on the availability of the underlying NWP model, but can be anywhere from hourly out to 18 h for contiguous U.S. (CONUS) applications, and at 3-h intervals out to 36 h for global applications. In Part II of this paper (Pearson and Sharman 2017), we will describe a nowcast application that uses a short-term forecast nudged to get better agreement with the most recent turbulence observations. The nowcast is updated at 15-min intervals and is particularly useful for defining areas of turbulence associated with convection.

2. Turbulence diagnostics

Atmospheric turbulence that affects aircraft as “bumpiness” is most pronounced when the size of the turbulent eddies encountered is about the size of the aircraft; for commercial aircraft this would correspond

to eddy dimensions of ~ 100 m and would be even smaller for general aviation aircraft. It is impossible to directly and routinely forecast atmospheric motion at these scales, now or even in the foreseeable future. However, the larger (synoptic and meso-) scales of atmospheric motion are easily resolved by grid spacings (\sim tens of kilometers) used by most current operational NWP models. In general, at these large grid spacings, the subgrid-scale turbulence parameterizations cannot be expected to perform very well and, in fact, this is the case, as will be shown later. Linkages between large-scale atmospheric features and aircraft-scale turbulence must then be developed and tested, either through empiricism or the application of theoretically based principles, assuming a downscale cascade from the larger resolved scales to the aircraft scales (e.g., Lindborg 1999). These diagnostics of turbulence, which may be manually derived or computed automatically from NWP model output, depend in part on the source of the turbulence. Hence, different diagnostics may be used for MWT, CAT, and CIT (e.g., Fahey et al. 2016; Knox et al. 2016). In fact, using the ensemble mean of a suite of diagnostics seems to provide better performance than any one single diagnostic; this is the GTG approach (e.g., Sharman et al. 2006; Kim et al. 2011). A comprehensive list of automated turbulence diagnostics used in the previous version of GTG was provided in Sharman et al. (2006). A few other diagnostics have been added to this suite since then, most notably those developed by McCann et al. (2012) and Schumann (2012). An inventory of the current suite of automated turbulence diagnostics available in GTG is provided in Table 1. Most of these diagnostics were developed with the aim of predicting clear-air turbulence, but since they actually identify strong spatial gradients regardless of the modeled source, they are in fact more general. Thus the term CAT diagnostic is used rather loosely here and really indicates any diagnostic that successfully identifies large spatial gradients in the output NWP model fields, either in cloud or out of cloud, and so includes other sources besides classical Kelvin–Helmholtz instabilities, most notably convective sources. Numerically, most of the diagnostics are computed on constant geometric altitude (z) surfaces (with the exception of some that are computed on isentropic surfaces; see Table 1), using second-order centered horizontal differences including map-scale factors appropriate for the NWP map projection used (e.g., Haltiner and Williams 1983), and using first-order irregular vertical differencing in the native vertical coordinates (e.g., Anderson et al. 1984). Note that some CAT diagnostics in Table 1 depend on the thermal wind relation and, therefore, are not appropriate for use in equatorial regions. Still others depend on the tropopause height, which regardless of the definition used,

can quite often fail. A procedure that computes the average of tropopause heights using different definitions of the tropopause is given in the appendix.

One modification to some of the diagnostics that seems to provide better statistical performance is to divide the computed raw diagnostic value by a local modified Richardson number $Ri^* = \max(Ri, 10^{-3})$. Physically, this could be due to the diagnostics sensing NWP-model-resolved inertia-gravity waves, which when entering a region of low background Ri are more likely to break down and become turbulent. This effect has in fact been demonstrated in numerical simulations (Lane et al. 2004) and has been suggested as an important component of unbalanced flow diagnostic performance (Knox et al. 2008).

a. Conversion of diagnostics to EDR

One of the more difficult challenges associated with turbulence forecasting is the specification of robust thresholds for light, moderate, and severe turbulence. The primary difficulty is that these qualitative turbulence intensity categories are not appropriate for forecasting purposes, since they are aircraft dependent. For the reasons mentioned above, EDR is the preferred atmospheric turbulence forecast metric. To provide EDR, the diagnostic values (which can be of various ranges of magnitudes and units) must somehow be mapped to a corresponding $\varepsilon^{1/3}$ value ($\text{m}^{2/3} \text{s}^{-1}$). These can then be compared with pilot reports (PIREPs) converted to EDR after accounting for aircraft type (Sharman et al. 2014) or compared directly with EDR estimates from in situ equipped aircraft for verification purposes.

Here, the basis for mapping raw diagnostic values to $\varepsilon^{1/3}$ depends on the probability density functions of observed turbulence intensity ε in the troposphere and lower stratosphere. Previous field campaigns have consistently shown that the distribution of turbulence intensity as indicated by ε or $\varepsilon^{1/3}$ has an approximately lognormal distribution in both the planetary boundary layer (PBL) and the free atmosphere [e.g., Nastrom and Gage (1985), Cho et al. (2003), although they show a bimodal distribution in the PBL, Frehlich and Sharman (2004a), and Frehlich et al. (2004)]. Many of these studies were based on long legs of commercial aircraft flights and therefore are relevant to the scales of motion that are resolvable by current-generation NWP models.

The lognormal distribution can be written as

$$\frac{1}{(2\pi\sigma_{\ln\varepsilon^{1/3}}^2)^{1/2}} \exp\left\{-\frac{[\ln(\varepsilon^{1/3}) - \langle\ln(\varepsilon^{1/3})\rangle]^2}{2\sigma_{\ln\varepsilon^{1/3}}^2}\right\}, \quad (1)$$

where

$$\sigma_{\ln\varepsilon^{1/3}}^2 = \langle [\ln(\varepsilon^{1/3}) - \langle\ln(\varepsilon^{1/3})\rangle]^2 \rangle$$

and the angle brackets indicate an ensemble average. Therefore, any turbulence diagnostic or linear combination of diagnostics that can be generated from NWP model output for forecasting turbulence should also have a log-normal distribution; for example, Cho et al. (2003) show an approximate lognormal distribution of shear, stability, and Richardson number in aircraft data. It should be noted that the lognormal distribution of $\varepsilon^{1/3}$ does have some theoretical difficulties (Frisch 1995); however, the experimental evidence does support its use within the current context. To produce a methodology that is consistent with the expected lognormal distribution of atmospheric turbulence intensity, a simple statistical mapping approach is used to transform a turbulence diagnostic D (which typically does not have turbulence intensity units and can vary over a range of several decades in these units) into $\varepsilon^{1/3}$.

Assuming lognormal properties then, the simplest mapping between a raw turbulence diagnostic D and $\varepsilon^{1/3}$ is provided by

$$\ln D^* = \ln(\varepsilon^{1/3}) = a + b \ln D, \quad (2)$$

where D^* is the $\varepsilon^{1/3}$ value corresponding to the raw diagnostic value D . The actual parameters for the mapped lognormal distribution are not that critical for designing a linear combination of diagnostics but it is desirable to use parameters that best describe the climatology. This provides constraints on a and b ; for example, applying the standard deviation (SD) operator to Eq. (2) gives

$$b = \frac{\text{SD } \ln \varepsilon^{1/3}}{\text{SD } \ln D} = \frac{C_2}{\text{SD } \ln D}. \quad (3)$$

Similarly, a is determined from the expectation operator $\langle \cdot \rangle$,

$$a = \langle \ln \varepsilon^{1/3} \rangle - b \langle \ln D \rangle = C_1 - b \langle \ln D \rangle. \quad (4)$$

Thus, the critical parameters in this mapping are the climatological values of the average and SD of $\ln \varepsilon^{1/3}$:

$$\begin{aligned} C_1 &= \langle \ln \varepsilon^{1/3} \rangle \quad \text{and} \\ C_2 &= \text{SD } \ln \varepsilon^{1/3}. \end{aligned} \quad (5)$$

These values are dependent on altitude and latitude at a minimum (e.g., Frehlich and Sharman 2010), and are difficult to derive because they require large numbers of unbiased samples to achieve robust results. Previous estimates of C_1 and C_2 , both from field programs and from lognormal fits to climatologies of in situ EDR data, were given in Sharman et al. (2014) for upper levels. Using that same fitting technique on a larger

TABLE 1. Definitions of CAT and MWT diagnostics.

Diagnostic	Description	Unit	Reference
Speed	Horizontal wind speed	m s^{-1}	Endlich (1964)
VWS	Vertical shear of horizontal wind	s^{-1}	Endlich (1964)
1/Ri	Inverse Ri (dry)		Sharman et al. (2006)
1/Ri _s	Inverse Ri (moist)		Sharman et al. (2006)
1/RiT _W	Inverse Ri with VWS from thermal wind relation		Sharman et al. (2006)
HS	Horizontal shear of horizontal wind	s^{-1}	Dutton (1980)
Endlich	Wind speed \times turning	radian s^{-1}	Endlich (1964)
Dutton	Dutton's empirical index	s^{-1}	Dutton (1980)
wsq	Vertical velocity (w) squared	$\text{m}^2 \text{s}^{-2}$	
iawind	Inertial advective wind	m s^{-2}	
DIV	Horizontal divergence	s^{-1}	
VORTSQ	Vertical vorticity ²	s^{-2}	
DEFSQ	Total deformation ²	s^{-2}	
-DRiDt	Ri tendency	s^{-1}	Roach (1970)
Brown1	Simplified DRi/Dt	s^{-1}	Brown (1973)
Brown2	$\epsilon^{1/3}$ from simplified DRi/Dt	$\text{m}^{2/3} \text{s}^{-1}$	Brown (1973)
EDRLUN	$\epsilon^{1/3}$ from simplified DRi/Dt	$\text{m}^{2/3} \text{s}^{-1}$	Gill and Buchanan (2014)
CP	Colson–Panofsky TKE	$\text{m}^2 \text{s}^{-2}$	Colson and Panofsky (1965)
Ellrod1	TI1 index	s^{-2}	Ellrod and Knapp (1992)
Ellrod2	TI2 index (includes divergence)	s^{-2}	Ellrod and Knapp (1992)
Ellrod3	Ellrod2 with divergence tendency substituted for divergence and divergence tendency computed as in Sharman et al. (2006)	s^{-2}	Ellrod and Knox (2010)
SGSTKE	Subgrid TKE from NWP model	$\text{m}^2 \text{s}^{-2}$	
DTF3	Diagnostic TKE–epsilon formulation	$\text{m}^2 \text{s}^{-2}$	Marroquin (1998)
LAZ	TKE	$\text{m}^2 \text{s}^{-2}$	Laikthman and Al'ter-Zalik (1966)
SEDR	$\epsilon^{1/3}$ derived from TKE	$\text{m}^{2/3} \text{s}^{-1}$	Schumann (2012)
SCHGW	$\epsilon^{1/3}$ derived from variance of vertical velocity	$\text{m}^{2/3} \text{s}^{-1}$	Schumann (2012)
NCSU1	Advection \times vorticity gradient /Ri*	s^{-3}	Kaplan et al. (2004)
NCSU2	Vorticity gradient \times gradient of Montgomery streamfunction computed on isentropic surfaces	s^{-3}	Kaplan et al. (2004)
PVGRAD	Horizontal potential vorticity gradient	PVU km^{-1}	
TROPG/Z	Tropopause gradient/tropopause height	m^{-1}	
NVA	Negative vorticity advection	s^{-2}	Sharman et al. (2006)
Stone	Inertial instability index	s^{-2}	Stone (1966); Knox (1997)
AGI	Anomalous gradient instability	s^{-2}	Alaka (1961); Mogil and Holle (1972)
LHFK	Lighthill–Ford spontaneous imbalance	m	Knox et al. (2008)
TEMPG	Horizontal temperature gradient	K m^{-1}	Sharman et al. (2006)
NGM1	Wind speed \times deformation	m s^{-2}	Reap (1996)
NGM2	Vertical temperature gradient \times deformation	$\text{K m}^{-1} \text{s}^{-1}$	Reap (1996)
F2D	2D frontogenesis function on constant z surfaces	K m s^{-1}	Sharman et al. (2006)
F2DTW	2D frontogenesis function on constant z surface using thermal wind relation	K m s^{-1}	Sharman et al. (2006)
F3D	3D frontogenesis function on constant z surfaces	K m s^{-1}	
FTH	Normalized 2D frontogenesis function computed on isentropic surfaces	$\text{m}^2 \text{s}^{-3}$	Sharman et al. (2006)
EDRLL	$\epsilon^{2/3}$ estimated from second-order longitudinal structure function	$\text{m}^{4/3} \text{s}^{-2}$	Frehlich and Sharman (2004a)
EDRAVG	$\epsilon^{2/3}$ estimated from average of longitudinal and transverse second-order structure functions	$\text{m}^{4/3} \text{s}^{-2}$	Frehlich and Sharman (2004a)
SIGWX	Variance of w estimated from second-order longitudinal structure function	$\text{m}^2 \text{s}^{-2}$	Frehlich and Sharman (2004b)
SIGWAVG	Variance of w estimated from average of longitudinal and transverse second-order structure functions	$\text{m}^2 \text{s}^{-2}$	Frehlich and Sharman (2004b)
CTSQ	Temperature structure constant estimated from average of longitudinal and transverse second-order structure functions of T	$\text{K}^2 \text{m}^{-2/3}$	Frehlich et al. (2010)

TABLE 1. (*Continued*)

Diagnostic	Description	Unit	Reference
MWT1	$d_s \times \text{WSQ/Ri}^*$	$\text{m}^3 \text{s}^{-3}$	
MWT2	$d_s \times \text{CTSQ}$	$\text{K}^2 \text{m}^{1/3} \text{s}^{-1}$	
MWT3	$d_s \times \text{F3D}$	$\text{K m}^2 \text{s}^{-2}$	
MWT4	$d_s \times \text{speed}$	$\text{m}^2 \text{s}^{-2}$	
MWT5	$d_s \times \text{DIV} $	m s^{-2}	
MWT6	$d_s \times \text{NGM1}$	m s^{-2}	
MWT7	$d_s \times \text{SIGWAVG}$	$\text{m}^3 \text{s}^{-3}$	
MWT8	$d_s \times \text{LHFK/Ri}^*$	$\text{m}^2 \text{s}^{-1}$	
MWT9	$d_s \times \text{iawind}$	$\text{m}^3 \text{s}^{-2}$	
MWT10	$d_s \times \text{EDRAVG}$	$\text{m}^{7/3} \text{s}^{-3}$	
MWT11	$d_s \times \text{SCHGW}$	$\text{m}^{5/3} \text{s}^{-2}$	
MWT12	$d_s \times \text{TEMPG} $	K s^{-1}	
MWT13	$d_s \times \text{DEFSQ}$	m s^{-2}	
MWT14	$d_s \times \text{EDRLL}$	$\text{m}^{7/3} \text{s}^{-3}$	

sample of data from selected Delta Air Lines (DAL) Boeing 737 (B737) and B767 and Southwest Airlines (SWA) B737 in situ equipped EDR aircraft for the time period 2009–14 for different altitude bands gives the values of C_1 and C_2 in Table 2. Note the values of both C_1 and C_2 are not strongly dependent on altitude, and for this reason the vertical average value (right column) was used. Besides its simplicity, this statistical mapping approach using climatologies of the peak (over 1 min) reports has the advantage of implicitly accounting for the fact that the diagnostics values are gridcell averages.

To actually map the raw diagnostic value D into $\epsilon^{1/3}$, each diagnostic is fit to the assumed lognormal distribution Eq. (1) by the following three-step procedure. First, a histogram of the distribution of each D is computed by binning all computed values of D derived once daily over 1 yr from the input NWP model forecast data for lead times of most operational interest (typically 6 h for U.S. operations and 12 h globally). One year provides a sample over a wide range of seasonal/synoptic conditions. The histogram values are normalized by the total number of data points in all bins multiplied by the bin width. Next, the histogram is fit to the lognormal distribution in Eq. (1) to obtain $\ln \epsilon^{1/3}$ and $\sigma_{\ln \epsilon^{1/3}}^2$. These two fit parameters are obtained using Powell's method (Press et al. 1986) to

minimize the relative mean square error e of the bin summation difference

$$e = \sum_{k=k_1}^{k_2} [\ln(H_k) - \ln(P_k)]^2, \quad (6)$$

where H_k is the probability density for diagnostic D at bin k and P_k is the lognormal density function [Eq. (1)] corresponding to the center of the k th bin. The sum is taken over the bins to fit, from $k = k_1$ to $k = k_2$, with the number of samples in each bin required to be >1000 . Typically, 50 bins were used, with $k_1 = 10$ and $k_2 = 49$. The small values of the magnitude of D contained in the first few bins are not of practical interest since they would correspond to very smooth nonturbulent conditions. Similarly, negative values of any index are not used in the fit. In any case, once the fit is performed and it is found that it would apply to lower or higher bins, the data can be refit to include them, or if the fit does not seem to apply to some bins in this range, the fitting range could be adjusted accordingly. Finally, for C_1 and C_2 in Table 2, the fit parameters b and a are obtained from Eqs. (3) and (4), respectively.

Figure 1 shows a lognormal fit to the distribution of the commonly used Ellrod1 index [the product of vertical wind shear times the total deformation; Ellrod and Knapp (1992)] computed from 1 yr of WRF Rapid

TABLE 2. Values of the coefficients C_1 and C_2 in Eq. (5) using DAL B737 and B767 and SWA B737 in situ EDR data for 2009–14. The number of observations is n_{obs} .

Climatological value	Altitude region (number of observations)			
	0–10 000 ft ($n_{\text{obs}} = 2906853$)	10 000–20 000 ft ($n_{\text{obs}} = 1020878$)	20 000–45 000 ft ($n_{\text{obs}} = 6517033$)	>0 ft ($n_{\text{obs}} = 10220469$)
C_1	−2.248	−2.578	−2.953	−2.572
C_2	0.4235	0.557	0.602	0.5067

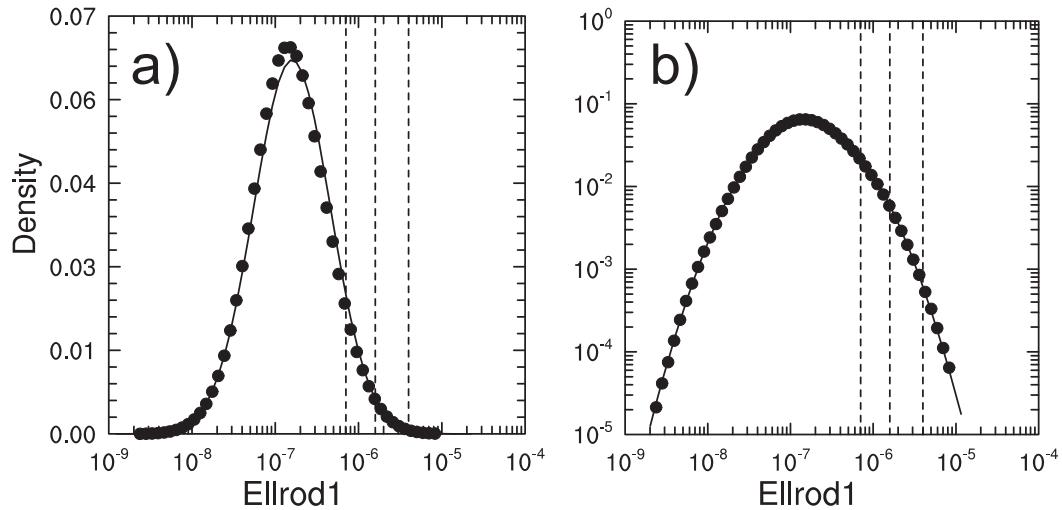


FIG. 1. Density plots of the Ellrod1 (Ellrod and Knapp 1992) turbulence diagnostic (s^{-2}) for upper levels (FL180–FL500) based on 6-h WRF-RAP NWP model forecasts over 1 yr (2015) in (a) log–linear and (b) log–log coordinates. Raw diagnostic values are binned into 50 bins indicated by the filled circles. The curve represents the best lognormal fit [Eq. (1)] of the diagnostic. Vertical dashed lines correspond from left to right to values of the diagnostic for which $\varepsilon^{1/3} = 0.15, 0.22, \text{ and } 0.34 m^{2/3} s^{-1}$.

Refresh (WRF-RAP) NWP model (Benjamin et al. 2016) data, with plots displayed in both log–linear and log–log coordinates. The log–linear plot in Fig. 1a shows the familiar lognormal distribution with the diagnostic distribution slightly skewed from the expected distribution defined by Eq. (1). The vertical dashed lines in the figure indicate values of the diagnostic that map to $\varepsilon^{1/3}$ values of $0.15, 0.22, \text{ and } 0.34 m^{2/3} s^{-1}$ (left to right, respectively). These values roughly correspond to light, moderate, and severe turbulence for medium-sized aircraft at cruise altitudes (Sharman et al. 2014) and lie considerably to the right of the peak in the distribution, and therefore the log–log version of the plot in Fig. 1b can be used to better assess the goodness of the fit in this region, which is of most practical interest. Other example diagnostic distributions and their lognormal fits are shown in Fig. 2 for WRF-RAP NWP-based diagnostics and in Fig. 3 for NOAA’s Global Forecast System (GFS) NWP model (<https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>) at 0.25° horizontal resolution and native hybrid-coordinates-based diagnostics. As can be seen for both the WRF-RAP and GFS NWP-based diagnostics, some fits are very close to the expected lognormal distribution over the entire range of D , while others do not fit well for very small or very large values of D . In any case, the critical range of D to fit is between the vertical dashed lines, and in that region all fits shown are quite good for all altitude bands. It should be noted that some diagnostics do not have the expected lognormal (not shown) shape, and if a robust lognormal

fit cannot be obtained for a particular diagnostic, it is not used. Although it cannot be stated that this diagnostic to EDR mapping procedure applies universally to all NWP models, the fact that it performs well for two different models (one nonhydrostatic regional and another hydrostatic global) with different resolutions, numerics, and physical parameterizations implies that it should also be successful for most other regional and global models.

b. Mountain-wave-turbulence diagnostics

Although conditions favoring large-amplitude mountain waves, overturning, and MWT are fairly well understood theoretically (for reviews see, e.g., Sharman et al. 2012; Wurtele et al. 1996; Fritts and Alexander 2003), forecasting their occurrence is still problematic, given their dependence on environmental conditions (which must be correctly forecast), nonlinear small-scale processes that are difficult to model even with fairly high-resolution simulation models (e.g., Doyle et al. 2011), and the limitations of the current 1D PBL parameterizations in complex terrain (Muñoz-Esparza et al. 2016). Operational NWP models are still fairly coarse ($\Delta x = \text{tens of kilometers}$) relative to the wavelength of internal-gravity waves and turbulent eddy sizes relevant for aircraft bumpiness. Therefore, current and previous methods used for forecasting MWT have been somewhat empirically based employing 1) predicted surface conditions (e.g., Calabrese 1966; Nicholls 1973; Hopkins 1977; Lee et al. 1984; Fahey et al. 2002), 2) parameterized wave stress

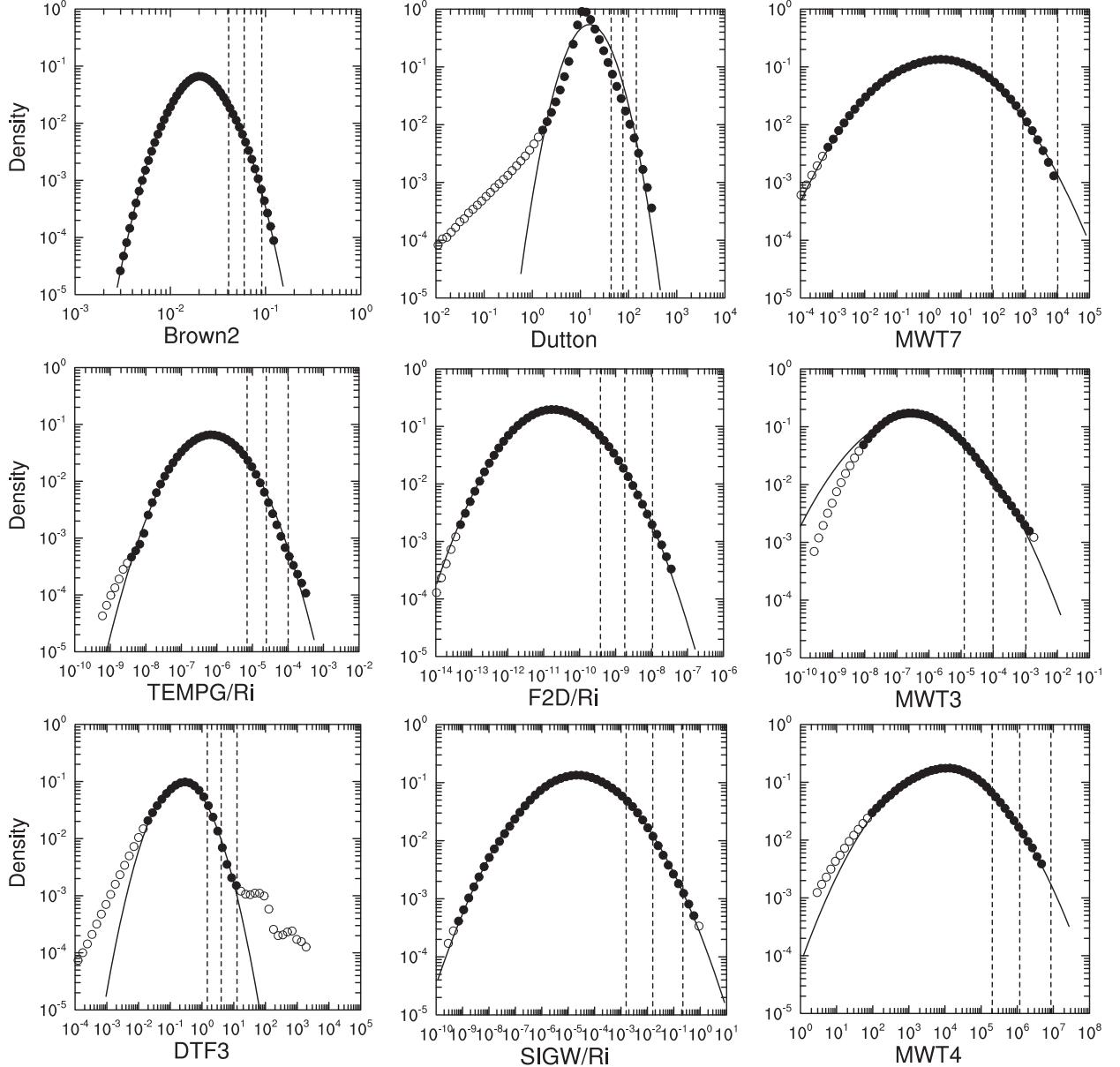


FIG. 2. As in Fig. 1b, except that sample distributions of various lognormal fits to some other turbulence diagnostics are shown for (top) upper levels, (middle) midlevels, and (bottom) low levels. The curve represents the best lognormal fit [Eq. (1)] of the diagnostics in the regions indicated by the filled circles. The open circles are not used in the fit. The diagnostics and their units are defined in Table 1.

profiles and gravity wave drag parameterizations (e.g., Kim and Arakawa 1995; Turner 1999; Kim and Chun 2011), 3) solutions to the linear steady-state vertical structure equation (e.g., Vosper 2003; Shutts 1997), or 4) ray-tracing methods (Eckermann et al. 2004; Eckermann et al. 2006). All of these approaches have generated some useful results but all still suffer from deficiencies. Some of these are outlined in Kim et al. (2003), but besides that, rigorous statistical evaluation of the predictive skill for forecasting MWT using any of these methods has not been performed.

Here, we take a somewhat pragmatic approach and use combinations of low-level wind speed and CAT diagnostics to produce 3D MWT diagnostics, which are then converted to EDR as described above. The use of low-level wind speed (the component perpendicular to the ridgeline) has had some success in the operational prediction of MWT; however, the result is then basically a 2D field constant with height, whereas observations often show substantial variations with height. On the other hand, we have found that traditional CAT diagnostics, although 3D, tend to underpredict the severity of

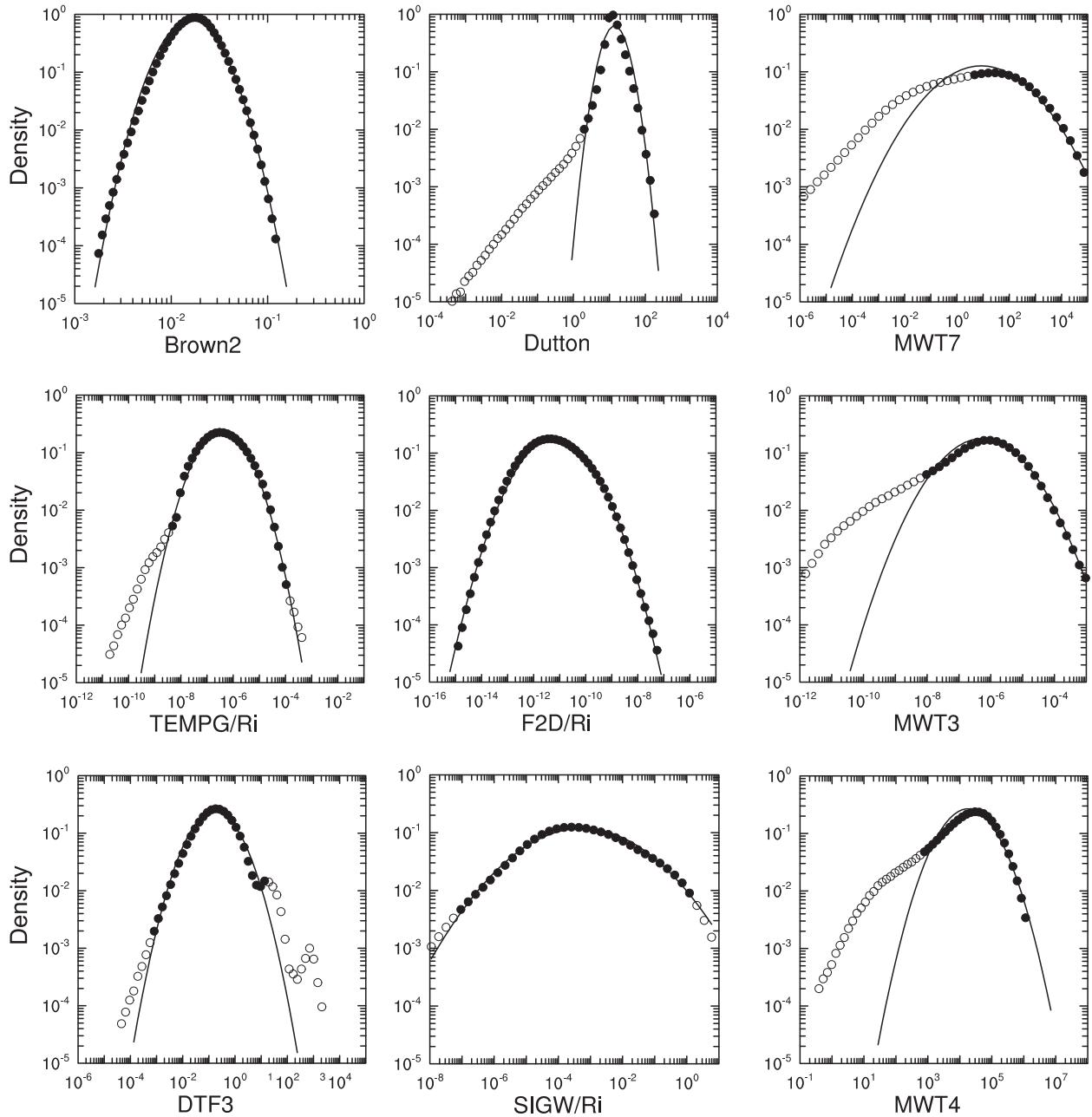


FIG. 3. As in Fig. 2, except that the sample distributions for various turbulence diagnostics and the lognormal fits are based on 12-h forecasts of the GFS NWP global model at 0.25° horizontal resolution and the native vertical structure.

MWT. Thus, it seems reasonable to use a combination of 2D MWT-related surface parameters and 3D CAT diagnostics that capture the larger-scale spatial variability:

$$D_{\text{MWT}} = d_s D_{\text{CAT}}, \quad (7)$$

where d_s is the near-surface diagnostic and D_{CAT} is a traditional 3D CAT diagnostic. Almost any

combination of surface wind speed, wind direction, imposed vertical velocity, terrain height, terrain height variance, or other terrain statistics could be used for d_s (Kim and Arakawa 1995; Kim and Doyle 2005), but after much experimentation the low-level wind speed V_s in conjunction with the NWP model terrain height h and terrain height gradient seemed to provide the best results. Namely, for each model grid point (i, j) ,

$$d_s = 0 \quad \text{if } h_{ij} < 200 \text{ m}, \quad \text{grad}(h_{ij}) < 5.0 \text{ m km}^{-1} \quad \text{and} \quad (8a)$$

$$d_s = V_s(\max \text{ in lowest } 1500 \text{ m}) \times \min(h_{ij}, 2750 \text{ m}) \text{ otherwise.} \quad (8b)$$

Note that since wind speed is used, the wind direction does not enter into this formulation, so determination of ridgelines as used in some other algorithms (e.g., Bacmeister et al. 1994; Eckermann et al. 2004) or identification of mountain-wave boxes/activating wind directions (e.g., Calabrese 1966; Lee et al. 1984; Fahey 1993) is not required. This is especially advantageous for global applications. The computed imposed vertical velocity at the lower boundary could be used in place of the wind speed in Eq. (8b), but evaluations of the statistical performance of this variant showed lower scores (by the performance metrics outlined in section 3 below).

After some experimentation, the 14 MWT diagnostics given in Table 1 were determined to provide the best overall discrimination performance. Each of these diagnostics can be mapped to EDR in an identical manner to that used for the CAT diagnostics. Figures 2 and 3 show some examples of the fits for the assumed lognormal distribution for MWT diagnostics.

c. Ensemble mean of diagnostics

As in other studies (Sharman et al. 2006; Kim et al. 2011; Gill and Stirling 2013; Gill and Buchanan 2014; Knox et al. 2016), the final turbulence forecast is provided as the ensemble mean of the turbulence diagnostics (GTG), each converted to EDR:

$$\text{GTG} = \frac{1}{N} \sum_{i=1}^N W_i D_i^*, \quad (9)$$

where N is the number of diagnostics used in the combination, typically 5–10. Here, D^* represents a particular turbulence diagnostic D (either CAT or MWT) remapped to an EDR scale ($\text{m}^{2/3} \text{s}^{-1}$), and W is an optional weighting of the diagnostic. The weights in Eq. (9) can be prescribed either dynamically or statically as described in Sharman et al. (2006). Short-term forecasts (0–2 h) use dynamic weights where each diagnostic performance is evaluated in real time based on agreement with the current observations (from available PIREPs and in situ EDR data); see Sharman et al. (2006) for details. Longer lead-time forecasts use static weights that are based on overall performance from a 1–2-yr retrospective period, or may simply be set to unity.

The suite of turbulence diagnostics chosen depends on the overall performance of each diagnostic

separately in each of three altitude bands: low levels (surface–10 000 ft MSL; 1 ft $\cong 30.5$ cm), middle levels (10 000 ft MSL–FL200 or approximately 20 000 ft MSL), and upper levels (FL200–FL500). The choice also depends on the desire to ensure that the diagnostics used are actually attempting to identify different atmospheric processes that may be contributing to turbulence (i.e., the diagnostics are uncorrelated with one another), as well as that the lognormal fits to the diagnostic look reasonable. The suite selected also depends on the underlying NWP model from which the diagnostics are computed. Optionally, once the GTG combination in Eq. (9) is computed in each altitude band, it could also be fit to the lognormal distribution in Eq. (2). The final output is based on an overlapping merger of the low-, middle-, and upper-level forecasts on the same horizontal grid structure as the NWP model. The procedure for computing the diagnostics is quite general and has been applied to many operational NWP models.

Two forecasts are provided. The first is a CAT forecast, which includes all sources of turbulence that have sufficient impact to the large scale to be detected by turbulence diagnostics through spatial inhomogeneities in the underlying NWP model. This may include large-scale convective sources as well as classical clear-air sources (e.g., deformation or shear). The second is an MWT-specific forecast. Both the CAT and the MWT forecasts are based on the ensemble mean of the diagnostics in Eq. (9) and are provided to users separately. A gridpoint-by-gridpoint maximum of the CAT and MWT forecasts is also provided. The set of diagnostics chosen for use in the combination depends on which statistical performance metric or combination of metrics is most suitable for a deterministic forecast. This is discussed in the next section. But the suite of diagnostics can also be used to provide an uncalibrated probability, as done in Krozel et al. (2011) and Kim et al. (2015). Examples of CAT and MWT ensemble mean GTG forecasts based on 6-h forecasts from the WRF-RAP NWP model are provided in Fig. 4. Figure 5 displays a GTG example based on 12-h GFS model output. These are fairly typical representations, and observations are overlaid to give some idea of the accuracy of these particular forecasts.

3. Evaluation of the diagnostics

The final step in evaluating the mapped diagnostics D^* is to compare the results with observations, which must also be in EDR units (i.e., $\text{m}^{2/3} \text{s}^{-1}$). For comparison with in situ EDR reports, this is straightforward; however, PIREPs are aircraft dependent, and therefore to convert the reported turbulence intensity (which is

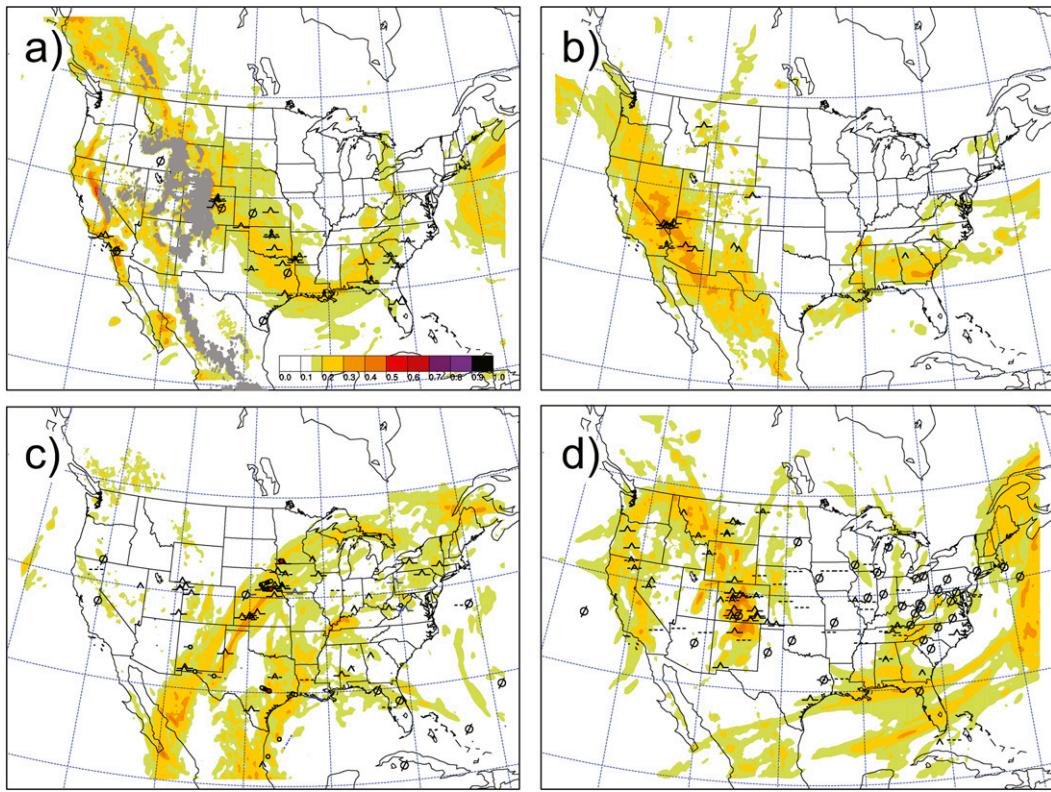


FIG. 4. Example GTG output based on 6-h WRF-RAP forecasts. Contours are EDR ($\text{m}^{2/3} \text{s}^{-1}$) color coded according to the color bar. Shown are the (a) CAT forecast valid at 1800 UTC 8 Feb 2016 at 7000 ft MSL, (b) CAT + MWT forecast valid at 2100 UTC 22 Feb 2016 at 12 000 ft MSL, (c) CAT forecast valid at 0000 UTC 10 Mar 2016 at FL390, and (d) CAT + MWT forecast valid at 1500 UTC 29 Jan 2016 at FL350. PIREPs within 1 h of the valid time using the standard symbols (e.g., <http://www.aviationweather.gov/adds/pireps/displaypireps>) and in situ EDR reports within 0.5 h of the valid time as color circles with the same color scale as the contours are overlaid for comparison. In (a), the gray areas represent intersecting terrain.

presumably a subjective measure of peak loads; Bass 1999) to a corresponding EDR requires information about the unique properties of the aircraft [e.g., aerodynamic chord, mass, and airspeed; see Cornman et al. (1995) and Cornman (2016)]. For most purposes it is sufficient to account for these effects in broad categories of aircraft weight class, such as are provided in the International Civil Aviation Organization (ICAO) definitions of “light,” “medium,” and “heavy” (http://www.faa.gov/air_traffic/publications/atpubs/CNT/5-1-1.htm). On the basis of comparisons of PIREP with in situ $\epsilon^{1/3}$ for medium-weight-class aircraft and the expected load dependence for different aircraft types, Sharman et al. (2014) derived the following relationship:

$$\begin{aligned} \epsilon^{1/3}(\text{peak over 1 min})_{\text{PIREP}} \\ = C \times R(W) P^2, \quad C = 0.0138, \end{aligned} \quad (10)$$

where $R(W)$ is an aircraft-weight-class-dependent factor given in Table 3 and P is the PIREP intensity category of

smooth to extreme converted to a 0–8 scale, with 0 corresponding to smooth, 1 to smooth-light, and 2, 4, 6, and 8 to light, moderate, severe, and extreme, respectively. The stated value of C is appropriate for upper levels but is probably altitude dependent.

Consistent with previous turbulence prediction assessments (e.g., Sharman et al. 2006; McCann et al. 2012;

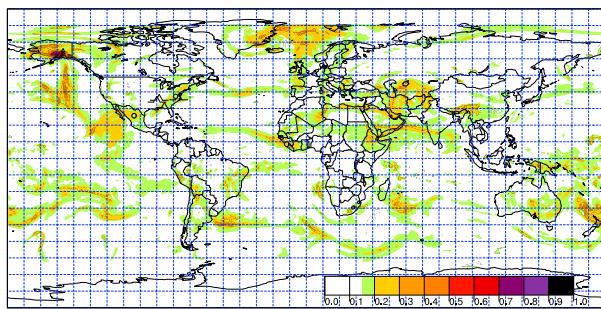


FIG. 5. As in Fig. 4, except that the GTG CAT + MWT forecast is based on a 12-h GFS forecast valid at 1800 UTC 30 Dec 2015 at FL330.

TABLE 3. The weight-class-dependent parameter R used in Eq. (10). The weight-class definitions are available online (http://www.faa.gov/air_traffic/publications/atpubs/CNT/5-1-1.htm).

ICAO weight class	Weight/mass limits (W)	$R(W)$
Light	$W \leq 15\,500 \text{ lb (7000 kg)}$	0.82
Medium	$15\,500 < W < 300\,000 \text{ lb}$ $(7000 < W < 136\,000 \text{ kg})$	1.0
Heavy	$\geq 300\,000 \text{ lb (136 000 kg)}$	1.22

Gill and Stirling 2013; Gill 2016), the overall discrimination ability for the individual diagnostics and ensemble mean of the diagnostics is obtained by computing probabilities of detection of turbulence areas above (yes) and below (no) what is normally considered to be a moderate turbulence intensity threshold. In this method, a contingency table of observations (PIREPs converted to EDR and in situ EDR estimates) in comparison with forecast turbulence diagnostic D^* values (using the grid point nearest to the observation) is formed for a given turbulence (EDR) observation threshold. From this table several different performance metrics may be derived (e.g., Jolliffe and Stephenson 2003; Gill 2016). Examples include the fraction of correctly forecast events (the hit rate or probability of detection, PODY, or simply POD) and the fraction of forecast events that were not observed (the false-alarm rate or probability of false detection, where POFD = $1 - \text{PODN}$). As shown by Brown and Young (2000), combinations of POD and POFD are preferable to the use of the false-alarm ratio (FAR) in assessing statistical performance since they are less susceptible to the relative frequencies of yes and no observations. Relative operating characteristic (ROC) curves can then be constructed for a given observation EDR threshold by applying continuous variations in forecast EDR threshold values to derive a curve of POD versus POFD values (e.g., Gill 2016). These ROC curves essentially measure the ability of a diagnostic to discriminate between yes and no observations. For the range of thresholds selected, higher values of both POD and POFD, and therefore larger areas under the POD–POFD curves (AUCs), imply greater discrimination skill. A no-skill discriminator will have an AUC ~ 0.5 , and a perfect discriminator will have an AUC = 1. Thus, the higher the AUC is in the range from 0.5 to 1, the better is the discrimination skill by this measure.

The statistical performance obviously depends on the threshold of interest. Typically, that threshold would be the light–moderate turbulence transition, since pilots would generally fly through what they consider light but would consider forecast moderate turbulence more

seriously. By comparing PIREPS of reported light, light–moderate, and moderate intensities with in situ EDR data from the same or nearby aircraft, Sharman et al. (2014, their Fig. 6) estimated that the median 1-min peak EDR for moderate reports was about $\sim 0.22 \text{ m}^{2/3} \text{ s}^{-1}$, and this threshold was used to compute contingency tables and consequent probabilities and ROC curves.

The diagnostics selected in the GTG combination in Eq. (9) also depend on which performance metric or metrics should be maximized or minimized for the GTG combination. Options include maximizing the AUC or partial AUC (i.e., the area under the ROC curve below some POFD threshold; e.g., Gill 2016), maximizing a combination of POD and POFD [such as the true skill statistic (TSS)], minimizing the fractional volume occupied by diagnostic values greater than the chosen threshold, minimizing the bias, minimizing the root-mean-square error (RMSE), or any combination of these. Several of these were tried, but overall, it seemed best to use the AUC, which has been traditionally used in other studies (e.g., Pepe and Thompson 2000; Marzban 2004; Sharman et al. 2006; Gill 2014). Forward selection of diagnostics starting with the one having the best AUC and adding diagnostics that increase the AUC until no further increase in AUC is obtained provides the set of diagnostics to use. During the selection process trial diagnostics that have a correlation coefficient > 0.85 with a previously selected diagnostic and diagnostics having an AUC < 0.7 are not used. This procedure is quite flexible and could easily be used considering other metrics or combinations of metrics in the evaluation process.

a. CAT diagnostics

Statistical evaluations were performed by computing ROC curves for the CAT diagnostics listed in Table 1 based on both the WRF-RAP NWP model 6-h forecasts and GFS NWP model 12-h forecasts. Observations used include PIREPs converted to EDR and routine 15-min and triggered peak DAL and SWA in situ EDR reports (Sharman et al. 2014) within ± 1 and ± 0.5 h, respectively, of the NWP valid times over a 1-yr period. No attempt is made to isolate and withhold observations that were related to convection, so all observations, regardless of the turbulence source, are used in these evaluations.

The evaluations of the CAT diagnostics derived from the WRF-RAP NWP model include those over the 1-yr time period starting 1 April 2015 for 6-h forecast valid times corresponding to those hours over the United States containing the highest air traffic density but that were not too close together to contain correlated physical processes (1500 and 2100 UTC). The spatial distribution

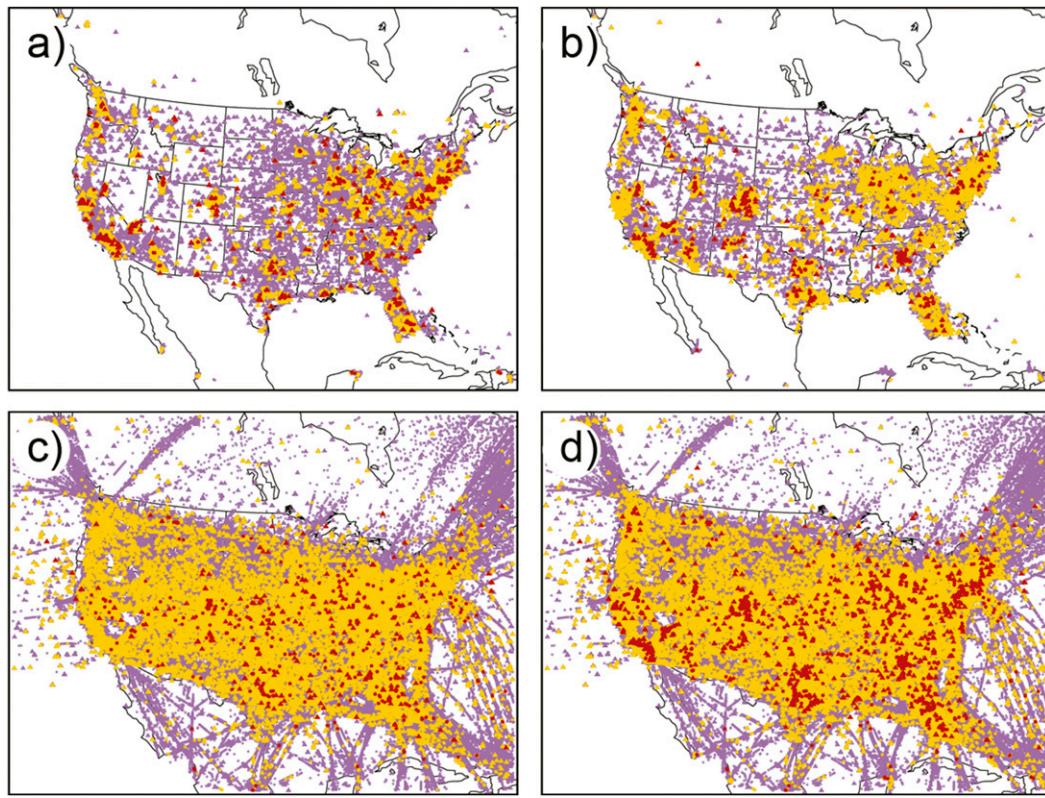


FIG. 6. Spatial distribution of observations used in the statistical evaluations of the WRF-RAP NWP-model-based turbulence diagnostics at (a) low levels, $n_{\text{obs}} = 91\,318$; (b) midlevels, $n_{\text{obs}} = 84\,910$; (c) upper levels, $n_{\text{obs}} = 693\,379$; and (d) all levels, $n_{\text{obs}} = 869\,607$. Colored triangles indicate PIREPs, and colored circles indicate in situ EDR reports. Reports of smooth values are colored purple, reports of moderate intensity are colored orange, and reports of severe are colored red ($\epsilon^{1/3} = 0.15, 0.22, \text{ and } 0.34 \text{ m}^{2/3} \text{ s}^{-1}$, respectively).

of the observational data used is provided in Fig. 6. As can be seen, the density of the observations at mid- and low levels is highest around the major airports of the United States, and especially over the eastern corridor, but at upper levels the density is more uniform over the entire contiguous United States.

Figure 7 shows the derived ROC curves for each of the three altitude bands for individual diagnostics, and for reference the Mellor–Yamada–Nakanishi–Niino (MYNN) parameterized subgrid-scale turbulent kinetic energy (SGS TKE; Nakanishi and Niino 2009) and the commonly used Ellrod1 index (TI1; Ellrod and Knapp 1992), along with the GTG combination in Eq. (9) using equal weights. The performance of most diagnostics as well as the GTG combination decreases from high to low altitudes, except for the NWP-model-produced SGS TKE, which is best at low levels. This latter characteristic is not surprising, since most SGS TKE parameterizations were developed for use in the PBL. In general, the Ellrod1 index is a very good discriminator at upper levels, but its success is relatively poor at mid- and low levels. Using the ensemble mean (GTG) provides benefit at all levels. The

statistical significance of the results has been evaluated by subsampling the observation data as in Sharman et al. (2006), and it was found that the minimum and maximum of the subsampled curves fit the average very tightly (not shown).

The list of CAT diagnostics used in the GTG CAT ensemble for each altitude band is provided in Table 4. Again, a CAT diagnostic within this context is any algorithm that senses large spatial gradients, including those caused by resolved convection or other sources. The list is based on those diagnostics that provided the maximum combined AUC, but could be different if another verification metric or combination of metrics or another NWP model is used or correlated diagnostics are allowed (Abernethy 2008; Guyon and Elisseeff 2003). The individual diagnostic and GTG ensemble mean AUC, PODY, PODN (=1 – POFD), TSS, and bias based on the 2×2 contingency table for an observation and forecast threshold of $\text{EDR} = 0.22 \text{ m}^{2/3} \text{ s}^{-1}$ and the lognormal fit coefficients a and b in Eq. (2) are provided. As can be seen, the diagnostics used in the GTG combination varies with altitude. Looking at the table, it is apparent the selection process trades off AUC

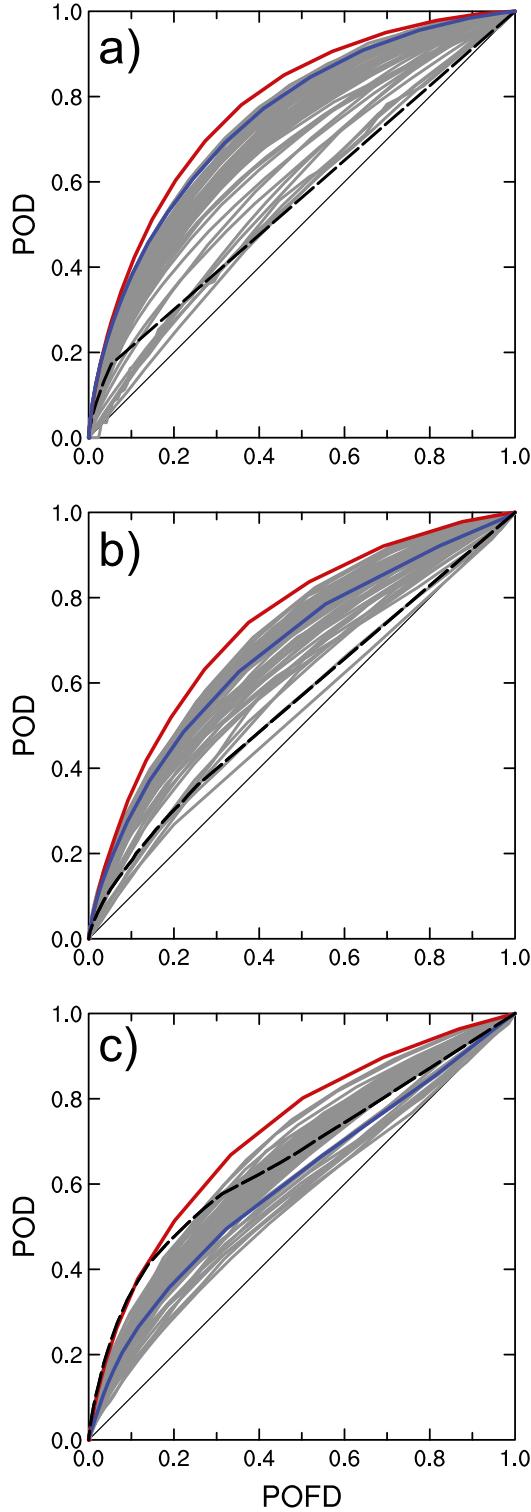


FIG. 7. ROC curves for CAT diagnostics (gray), GTG ensemble mean combination (red), and for reference the Ellrod1 index (blue), along with the MYNN SGS TKE (black dashed) in the (a) high-, (b) mid-, and (c) low-altitude bands. The diagonal line represents the no-skill line. Curves are constructed based on WRF-RAP 6-h forecasts with an observational threshold of $EDR = 0.22 \text{ m}^{2/3} \text{ s}^{-1}$. The number of observations used was 693 379 in (a), 84 910 in (b), and 91 318 in (c).

for TSS to achieve maximum AUC; that is, some diagnostics actually have a higher TSS than the TSS of the GTG combination, but according to our selection criterion, the AUC of the GTG combination is the highest. Note that for all diagnostics and the GTG combination the PODY and consequently TSS are relatively low (a perfect discriminator would have a TSS = +1), while the PODN is near 1. This is a reflection of the large number of negative reports (taken as smooth conditions with $EDR < 0.22 \text{ m}^{2/3} \text{ s}^{-1}$) relative to positive reports ($EDR \geq 0.22 \text{ m}^{2/3} \text{ s}^{-1}$), that is, the fact that elevated turbulence is a rare event. Of course, the PODY and TSS could be increased by lowering the threshold, but at the expense of a decrease in PODN and an increase in bias or volume of elevated turbulence forecast. Most biases are larger than 1, indicating some amount of overforecasting, but also this is due in part to the biases in the sampling of the observations (Brown and Young 2000). The difficulties associated with the scoring of rare events is not unique to turbulence, and other performance metrics based on the 2×2 contingency table have been suggested (e.g., Doswell et al. 1990), but in the end it really depends on what performance metric is most important to the end user.

EDR mappings, contingency tables, and ROC curves have also been developed using the GFS NWP model 12-h forecasts for the 1-yr period starting 1 January 2015 and valid at 1200 UTC as input. Not enough data were available for reliable statistical evaluations at mid- and low levels; however, the amount of data was sufficient at upper levels and Fig. 8 shows the spatial distribution. To determine if significant differences in the ROC curves exist regionally, two evaluations were performed: one using all available upper-level global observational data and another withholding the observational data over the CONUS (rectangle in Fig. 8). The results are shown in the ROC curves including all relevant observations (Fig. 9a) and excluding data over the CONUS (Fig. 9b); see also Table 5. The same criteria for determining the GTG combination that were used in the construction of the ROC curves for the WRF-RAP NWP-model-based diagnostics shown in Fig. 7 were also used for the GFS-based diagnostics. Again, the Ellrod1 index is a very good discriminator globally, but the GTG combination is better as measured by the AUC metric. By this metric the performance is about the same whether including or not the CONUS observations. However, the uncertainty in these results might be quite high as a result of the PIREP position error over the oceans.

b. MWT diagnostics

In evaluating MWT forecasts, a little more care must be exercised to ensure that the turbulence observations

TABLE 4. Indices used in the GTG CAT and MWT combinations, some of their respective statistical scores (AUC, and PODY, PODN, TSS, and bias computed from 2×2 contingency tables for an observation and forecast threshold of $\text{EDR} = 0.22 \text{ m}^{2/3} \text{ s}^{-1}$), and the fit coefficients a and b [Eq. (2)]. Results for the CAT forecasts are based on 1 yr (1 Apr 2015–31 Mar 2016) of WRF-RAP 6-h forecasts, while the MWT forecasts are based on the 6-yr time period 2010–15, both valid at 1500 and 2100 UTC. The numbers of observations used are given as n_{obs} .

Index	AUC	PODY	PODN	TSS	Bias	a	b
GTG CAT high ($n_{\text{obs}} = 693\,379$)							
Ellrod3	0.768	0.319	0.938	0.256	1.952	4.8428	0.47295
Fth/Ri	0.742	0.244	0.946	0.190	1.658	-0.032605	0.1639
NGM1	0.727	0.194	0.952	0.146	1.463	0.87721	0.4677
DEFSQ	0.744	0.220	0.956	0.177	1.819	5.275	0.3931
VORTSQ/Ri	0.718	0.182	0.963	0.145	1.139	3.8657	0.28582
iawind	0.705	0.222	0.932	0.154	2.007	-3.379	0.5577
EDRLL	0.708	0.179	0.956	0.134	1.342	0.62705	0.70497
1/RiTW	0.708	0.161	0.946	0.107	1.570	-1.508	0.4401
CTSQ/nRi	0.753	0.276	0.943	0.219	1.767	0.8727	0.24223
WSQ/Ri	0.734	0.345	0.893	0.238	3.141	-1.0263	0.15875
SIGWAVG/Ri	0.721	0.587	0.721	0.309	7.877	-0.7943	0.16044
GTG ensemble	0.788	0.221	0.965	0.186	1.131	0.5579	1.290
GTG CAT midlevel ($n_{\text{obs}} = 84\,910$)							
NGM1	0.690	0.191	0.943	0.134	0.588	1.086	0.4748
EDR	0.697	0.378	0.847	0.225	1.442	1.857	1.008
1/RiTW	0.701	0.384	0.845	0.229	1.525	-1.690	0.515
SIGWAVG/Ri	0.713	0.269	0.910	0.179	0.899	-0.994	0.175
F2DTW/Ri	0.695	0.219	0.938	0.157	0.652	3.738	0.2581
GTG ensemble	0.740	0.326	0.907	0.233	1.010	0.5370	1.325
GTG CAT low ($n_{\text{obs}} = 91\,318$)							
NGM1	0.664	0.133	0.966	0.099	0.545	1.255	0.4746
iawind	0.664	0.130	0.968	0.085	0.519	-2.920	0.4789
SIGWAVG/Ri	0.686	0.428	0.815	0.195	2.695	-1.092	0.1145
SGSTKE/Ri	0.658	0.440	0.807	0.248	2.801	-1.516	0.1238
GTG ensemble	0.721	0.174	0.963	0.137	0.633	0.6713	1.420
GTG MWT high ($n_{\text{obs}} = 649\,722$)							
MWT2	0.933	0.435	0.989	0.424	1.844	-2.508	0.2716
MWT12	0.932	0.399	0.991	0.390	1.270	0.44484	0.2434
GTG ensemble	0.933	0.435	0.989	0.424	1.844	-0.1419	0.9851
GTG MWT midlevel ($n_{\text{obs}} = 81\,147$)							
MWT9	0.820	0.324	0.951	0.275	15.935	-5.911	0.3138
MWT12	0.835	0.302	0.907	0.258	14.076	-1.857	0.3855
GTG ensemble	0.835	0.420	0.919	0.339	25.996	-0.006254	1.0196
GTG MWT low ($n_{\text{obs}} = 87\,113$)							
MWT4	0.768	0.043	0.980	0.023	16.904	-6.669	0.3683
MWT14	0.753	0.111	0.971	0.082	24.744	-2.845	0.2449
MWT6	0.780	0.182	0.959	0.140	36.445	-3.023	0.2808
MWT7	0.783	0.722	0.603	0.375	336.27	-3.560	0.2019
GTG ensemble	0.791	0.422	0.909	0.332	77.900	-0.0601	1.0122

used are related to mountain waves or at least occur over mountain-wave-prone regions. Fortunately, PIREPs sometimes contain text indicating a connection between wave motion and a turbulence experience. An example is

UA /OV SUN360035/TM 1837/FL125/TP PA31/ **TB MOD/RM MTN WAVE.**

Here, the OV gives position relative to the listed VHF Omni Directional Radio Range (VOR) transmitter, TM is the UTC time, FL is the flight level, TP is the aircraft type, TB is the subjective turbulence intensity, and the remarks section starts with RM. In this case the pilot is

reporting moderate turbulence associated with a mountain wave. As expected, some mountain-wave reports can indicate smooth turbulent conditions—for example,

UA /OV RLG/TM 1418/FL150/TP C172/WV 30050KT/ **TB NEG/RM TREMENDOUS MTN WAVE.**

Other pilots might not report a turbulence experience but do remark about the presence of mountain waves—for example,

UA /OV ALS 285030/TM 2105/ FL470/TP LJ45/**RM SEV MTN WAVE.**

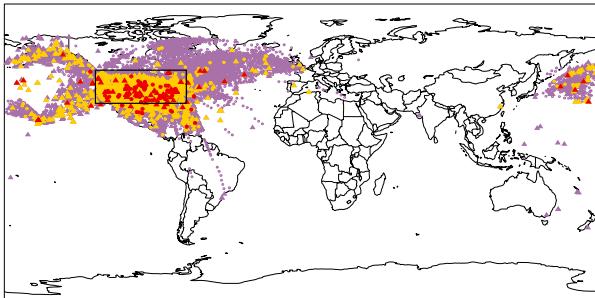


FIG. 8. As in Fig. 6, but displaying the spatial distribution of observations used in the statistical evaluations of the GFS NWP-model-based turbulence diagnostics at upper levels ($n_{\text{obs}} = 176\,062$). The black rectangle over the CONUS indicates observations withheld (137 369) in the non-CONUS statistical evaluations (see text).

Still others provide some qualitative assessment of wave amplitude—for example,

UUA /OV MVA 085050/TM 1835/FL400/TP B737/TB
SEV/RM SEV MTN WAVE/FULL TILT ON
THROTTLES. +/-40KTS.

As these examples show, turbulence and wave amplitude are often used interchangeably. There is no accepted definition of wave amplitude categories; however, in talking with pilots, it seems that larger-amplitude waves are as significant a hazard as the turbulence that may be associated with the “breaking” of the wave since it may force them off the assigned altitude and may lead to significant airspeed fluctuations that could ultimately stall the aircraft. Therefore, for the purposes of verification, the mountain-wave amplitude, if reported, is considered to be equivalent to a turbulence report of the same intensity. Thus, every PIREP that is missing a turbulence intensity but contains a “wave” or similar comment indicated in the PIREP remarks section is considered a turbulence report and is converted to EDR using Eq. (10). To include the information about waves in the scoring procedure required development of an automated PIREP decoder to parse the remarks section of the raw PIREPs.

The observations used to determine MWT diagnostic performance statistics include not only MWT-indicated PIREPs but also observations over MWT-prone areas. These areas can be defined in a number of ways, but here we simply used the definition provided by Eq. (8a) (i.e., $d_s > 0$). This is plotted in Fig. 10, using GFS 0.25° terrain data to define d_s , and nicely defines most known MWT-prone areas, including the western half of the United States; the Appalachian Mountains in the eastern United States; the southern tip of Greenland; the Andes; the Anchorage, Alaska, area; the Alps in Europe; and the

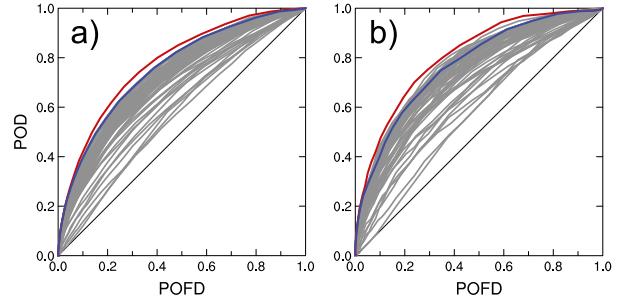


FIG. 9. As in Fig. 7, but for only upper levels and with curves based on the GFS 12-h forecasts. (a) All global observations ($n_{\text{obs}} = 176\,062$) and (b) only observations outside the CONUS ($n_{\text{obs}} = 7322$) are used.

Himalayas. Using these observational constraints, it was found that there were too few MWT-specific PIREPs to get robust statistics over one or even two years of data, so the dataset used for comparison was expanded to the 6-yr time period 2010–15, using 6-h WRF-RAP forecasts valid at 1500, 1800, 2100, and 0000 UTC. The time windows for comparison with the observations were the same as those used in the CAT diagnostic evaluations. The results are provided in the bottom rows in Table 4, as well as in Fig. 11. These statistics were generated in the same way as the CAT statistics but using only PIREPs, which explicitly indicate a mountain-wave-related event or null events (PIREPs and in situ EDR data) in areas where $d_s > 0$. Note the very large biases indicating substantial overforecasting, especially at the mid- and low levels, but this is in part due to the sampling bias introduced by restricting the results to MWT areas (Brown and Young 2000). Figure 11 shows the ROC curves generated in the same way as those in Fig. 7, but using only PIREPs, which explicitly indicate a mountain-wave-related event or null events (PIREPs and in situ EDR data) in areas where $d_s > 0$. The AUCs for the MWT diagnostics are now much higher than for the CAT diagnostics, and in particular are much better than the Ellrod1 index, or the commonly used gravity wave drag formulation (Palmer et al. 1986). Table 4 shows some statistics and the benefits of using an ensemble mean of diagnostics, but the results for mid- and low levels contain quite a bit of uncertainty because of the low number of positive reports, which also accounts for the jagged appearance of the ROC curves at mid- and low levels. Too few MWT-specific PIREPs were available outside the CONUS to provide reliable performance metrics globally. More observations would help, but for other parts of the world, the only strategy that could be used to develop and evaluate MWT diagnostics would be to focus on those observations over what are expected to be MWT-prone areas (Fig. 10).

TABLE 5. As in Table 4, but based on 1 yr (2015) of GFS-based CAT 12-h forecasts valid at 1800 UTC, for CAT at upper levels, including all global observations and excluding observations over the CONUS.

GFS index	AUC	PODY	PODN	TSS	Bias	<i>a</i>	<i>b</i>
GTG CAT high ($n_{\text{obs}} = 176\,062$)							
Ellrod3	0.759	0.308	0.937	0.245	2.316	4.884	0.4594
Fth/Ri	0.728	0.265	0.933	0.198	2.404	0.104	0.1743
DEFSQ	0.716	0.239	0.944	0.183	2.016	4.804	0.3528
iawind	0.689	0.282	0.894	0.174	3.725	-2.876	0.4872
UBF/Ri	0.724	0.223	0.952	0.175	1.757	-1.521	0.2889
PVGRAD	0.630	0.046	0.983	0.029	0.595	0.0686	0.4704
EDRLL	0.688	0.265	0.906	0.171	3.259	0.7403	0.6390
TEMPG/Ri	0.730	0.249	0.945	0.194	1.989	2.426	0.3304
NCSU2/Ri	0.739	0.125	0.966	0.092	1.203	2.250	0.1502
EDRLUN	0.685	0.299	0.917	0.216	2.958	1.182	0.7014
wsq/Ri	0.723	0.239	0.938	0.176	2.233	-0.195	0.1981
GTG ensemble	0.766	0.246	0.957	0.203	1.611	0.4727	1.215
GTG CAT high no CONUS ($n_{\text{obs}} = 38\,693$)							
Ellrod3	0.774	0.311	0.942	0.252	2.757	4.884	0.4594
NGM1	0.736	0.200	0.952	0.152	2.214	0.695	0.4074
VWS	0.729	0.270	0.948	0.218	2.445	2.319	0.8474
EDR	0.741	0.288	0.944	0.191	2.599	0.877	0.7343
DIV /Ri	0.781	0.303	0.946	0.249	2.550	1.884	0.3015
SIGWAVG/Ri	0.765	0.654	0.736	0.390	11.720	-0.2268	0.1791
GTG ensemble	0.808	0.346	0.947	0.293	2.567	0.5092	1.204

4. Summary and conclusions

A method has been developed to statistically map automated turbulence forecast diagnostics to the ICAO (2001) standard atmospheric turbulence metric of EDR. The method assumes a lognormal distribution of diagnostic values, and uses a climatological mean and standard deviation of $\ln \epsilon^{1/3}$ derived from 1-min peaks recorded from in situ EDR-equipped aircraft. Assuming the lognormal distribution, the techniques outlined in section 2a are used to fit the raw diagnostic histograms to obtain the two necessary parameters [*a* and *b* in Eq. (2)] for the diagnostic-EDR mappings. The fit parameters are obtained in a preprocessing step that needs to be performed only once for a given NWP model. However, fit parameters for other models will depend on the horizontal and vertical grid spacings of the model, since this would affect the magnitude of the computed spatial gradients. Regardless of the NWP model used, the climatological fit parameters C_1 and C_2 (Table 2) would be the same.

Explicit MWT diagnostics were developed and tested, and high discrimination skill was shown in comparison with CAT diagnostics when confining our interest to mountain-wave-prone areas. To do these evaluations, PIREPs had to be identified that specifically mentioned that the turbulence experienced was related to a wave. This required the development of an automated PIREP decoder to parse the remarks section of the raw PIREP.

Both CAT and MWT diagnostics were evaluated using the AUC, a standard statistical discrimination metric, although other metrics could be considered. No attempt was made to isolate and withhold observations that were due to convection; therefore, the actual skill in truly nonconvective conditions may be higher than that presented here. For both the CAT and MWT turbulence categories, using the ensemble-weighted mean of several diagnostics (GTG) provides higher AUCs than any one individual diagnostic. This was demonstrated for forecasts based on both the WRF-RAP and GFS NWP models. Similar results would be expected to hold if other input NWP models were used.

The procedure used to compute the simple ensemble mean could be substituted for other more complex combination strategies (e.g., logistic regression, random

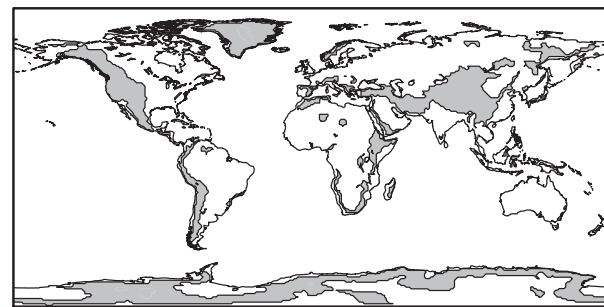


FIG. 10. MWT-prone areas as inferred from Eq. (8) applied to the GFS NWP model.

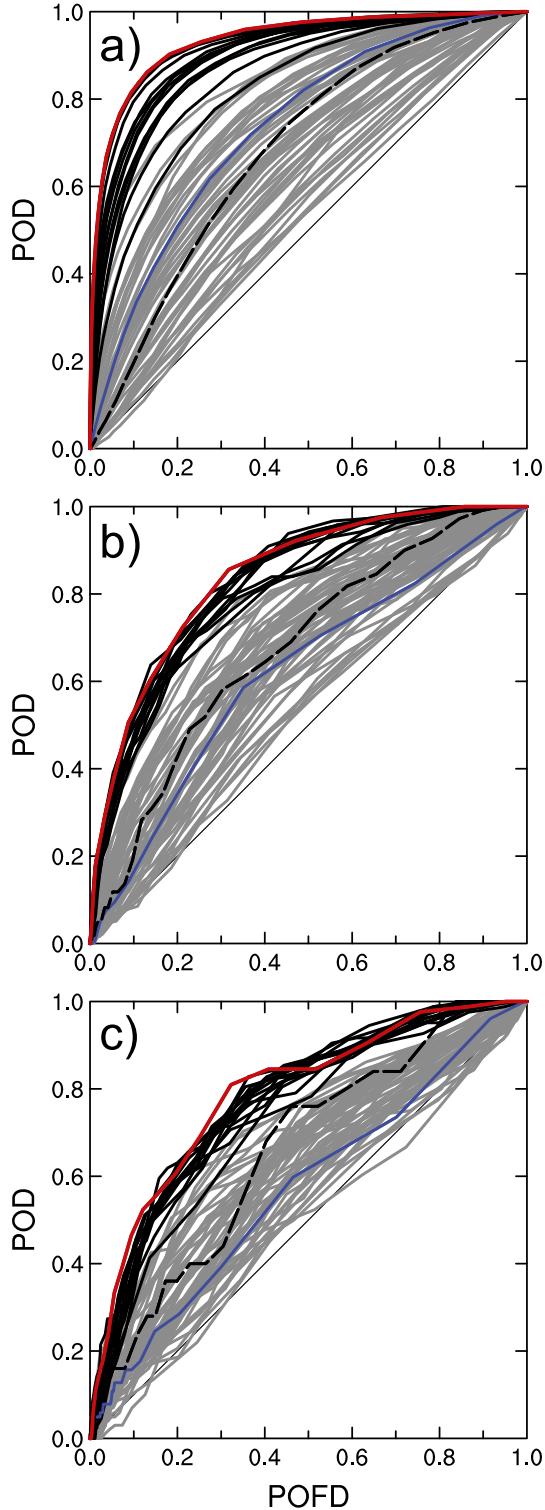


FIG. 11. As in Fig. 7, but for MWT diagnostics (black) and the GTG–MWT ensemble mean combination (red), and the black-dashed curve is the implementation of the traditional gravity wave drag algorithm (Palmer et al. 1986). The numbers of observations used were (a) 649 722, (b) 81 147, and (c) 87 113.

forecast, or support vector machines) and these have been tried in the past (Sharman et al. 2006; Abernethy 2008; Williams 2014), but given that the ROC curves were substantially elevated when using the new MWT algorithms (Fig. 11), it seems likely that the development of better diagnostics generally would provide the largest improvements in discrimination skill. Also, it must be remembered that some of the turbulence forecast error is due to errors in the underlying NWP model, and if it turns out that this is a large contributor to the overall error, developing better diagnostics or combination algorithms could not be expected to be helpful.

In any case, the development of new diagnostics is limited by our understanding of turbulence processes, especially in the stably stratified shear flow characteristic of the free atmosphere above the PBL. Indeed, recent high-resolution simulations of turbulence encounters by Trier et al. (2012), Kim et al. (2014), Zovko-Rajak and Lane (2014), and Trier and Sharman (2016) point out the complexities of the physical processes influencing the onset of upper-level turbulence and its dependence on deep convection. Fairly high horizontal resolution (<3 km, sometimes <1 km) was required to capture these events, indicating the need for higher-resolution operational NWP models to ultimately be able to produce acceptably accurate turbulence forecasts. With the introduction of such models it is likely that the subgrid- and resolved-scale TKE may start to become a more reliable predictor. Currently, it is only useful at low levels in the boundary layer (Fig. 7).

Finally, the highly transient and localized nature of turbulence implies fundamentally low predictability and providing meaningful deterministic forecasts will remain difficult. Therefore, probabilistic forecasts created by using NWP ensembles or using the information contained in the ensemble of individual diagnostics, or both (e.g., Kim et al. 2015), or by other probabilistic forecasting methods (e.g., Williams 2014), will probably see more use in the future. Still, the low predictability of turbulence associated with convective sources is particularly problematic, and until NWP models are better at forecasting the timing, location, and intensity of convective storms, using short-term nowcasts for tactical avoidance is probably the only viable approach. This is the subject of Part II of this paper (Pearson and Sharman 2017).

Acknowledgments. The authors are grateful to Stan Trier, Domingo Muñoz-Esparza, Ulrich Schumann, and to the anonymous reviewers for their constructive comments that led to clarifications in the manuscript, and also to Lara Ziady for constructing some of the figures. We are especially indebted to Rod Frehlich (now deceased) who originally suggested the lognormal diagnostic to EDR

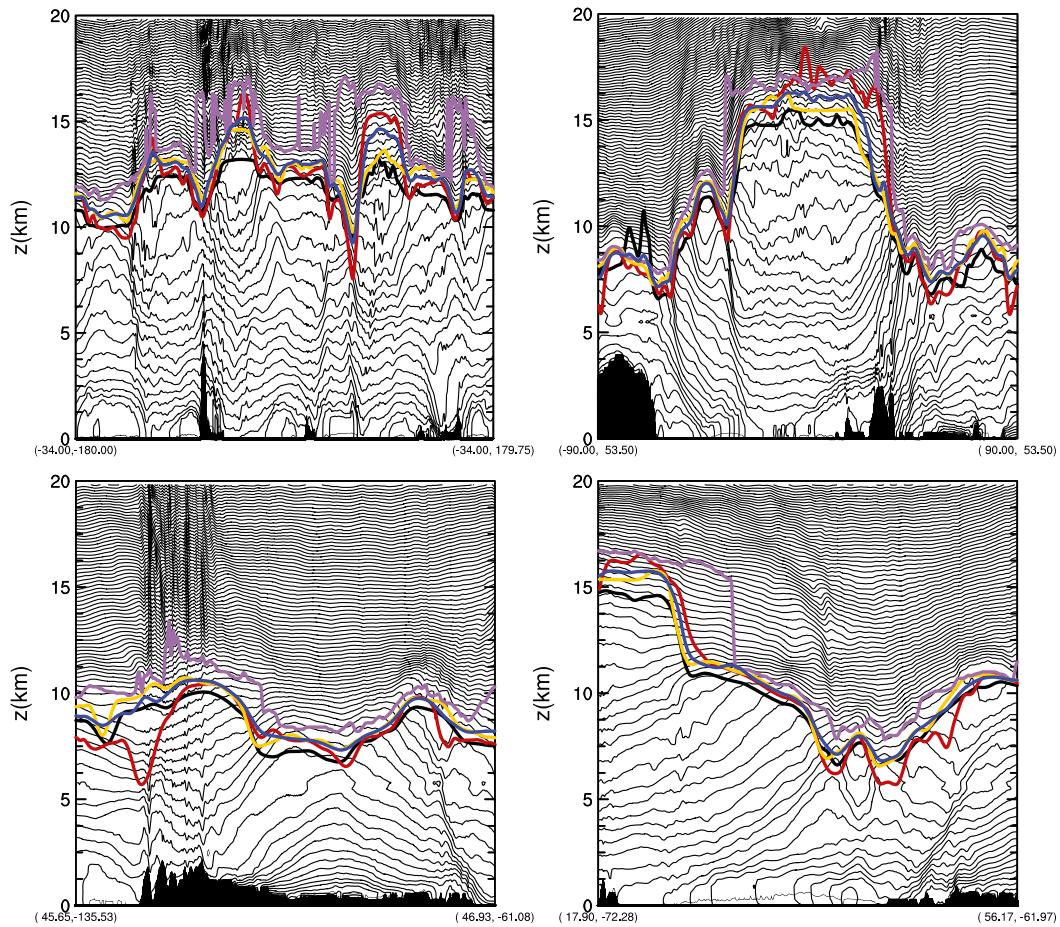


FIG. A1. Example cross sections through (top) GFS and (bottom) WRF-RAP model output. Shown are (left) east–west and (right) north–south cross sections. Thin black lines are isentropes. The lower boundary shows model terrain. The five heavily colored lines are the tropopause height computed from formulation 1 in magenta, 2 in black, 3 in red, and 4 in orange, with the average in blue.

statistical mapping approach. This research is in response to requirements and funding by the Federal Aviation Administration (FAA; Grant DTFAWA-15-D-00036). The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA.

APPENDIX

Computational Procedure for Tropopause Height

Some turbulence diagnostic formulations could depend either directly on the NWP-model-derived tropopause height or indirectly through a slightly different formulation in the stratosphere and troposphere. Thus, it becomes important to provide a robust calculation of the tropopause height. Several formulations were tried, and all were either biased or provided erratic results. Eventually, a formulation was used that involved the

average of several definitions. See Birner (2006) and Pan et al. (2004) and references therein for various tropopause definitions. The formulations that were used were

- 1) a lapse rate (γ) formulation that uses the standard WMO definition for the “first tropopause” of “the lowest level at which the lapse rate decreases to $2^{\circ}\text{C km}^{-1}$ or less, provided that the average lapse rate between this level and all higher levels within 2 km does not exceed $2^{\circ}\text{C km}^{-1}$,”
- 2) a lapse rate formulation that searches upward from the level where $p \sim 800\text{ hPa}$ to the vertical level where three consecutive grid points have $\gamma \leq 5^{\circ}\text{C km}^{-1}$ and specific humidity is less than one-tenth of its low-level ($>500\text{ hPa}$) average,
- 3) a potential vorticity (PV) formulation that searches downward from the model top to the vertical location where two consecutive grid points vertically averaged have $\overline{PV} < 2\text{ PVU}$ ($1\text{ PVU} = 10^{-6}\text{ m}^2\text{ K}^{-1}\text{ s}^{-1}$), and

- 4) a formulation that uses the ratio of a 4-km depth-averaged stability \bar{N}^2 between adjacent vertical grid points searching downward for the first point where the $\bar{N}_2^2/\bar{N}_1^2 > 1.1$ and $\bar{N}_2^2 \geq 1.7 \times 10^{-4} \text{ s}^{-2}$, where the subscripts 1 and 2 denote the lower and upper averages, respectively.

For each of these formulations, a median filter is used. Finally, an average of the four smoothed tropopause-height values is obtained using only those estimates that are within 3.5 km of one another. These were tested on both WRF-RAP and GFS NWP model output, and Fig. A1 shows some examples. Note that the first formulation (see the magenta lines in Fig. A1) can give very erratic results, which is caused by the low γ threshold. Similarly, the 2-PVU threshold sometimes provides tropopause-height estimates (red lines in Fig. A1) that are considerably lower than those estimates from the other definitions. In general, using the average of the four formulations gives much more robust results.

REFERENCES

- Abernethy, J. A., 2008: A domain analysis approach to clear-air turbulence forecasting using high-density in-situ measurements. Ph.D. dissertation, University of Colorado Boulder, 152 pp.
- Alaka, M. A., 1961: The occurrence of anomalous winds and their significance. *Mon. Wea. Rev.*, **89**, 482–494, doi:[10.1175/1520-0493\(1961\)089<0482:TOOAWA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1961)089<0482:TOOAWA>2.0.CO;2).
- Anderson, D. A., J. C. Tannehill, and R. H. Pletcher, 1984: *Computational Fluid Mechanics and Heat Transfer*. McGraw-Hill, 599 pp.
- Bacmeister, J. T., P. A. Newman, B. L. Gary, and K. R. Chan, 1994: An algorithm for forecasting mountain wave-related turbulence in the stratosphere. *Wea. Forecasting*, **9**, 241–253, doi:[10.1175/1520-0434\(1994\)009<0241:AAFFMW>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0241:AAFFMW>2.0.CO;2).
- Bass, E. J., 1999: Towards a pilot-centered turbulence assessment and monitoring system. *Proc. 18th Digital Avionics Systems Conf.*, St. Louis, MO, Institute of Electrical and Electronics Engineers, 6.D.3-1–6.D.3-8, doi:[10.1109/DASC.1999.821980](https://doi.org/10.1109/DASC.1999.821980).
- Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, doi:[10.1175/MWR-D-15-0242.1](https://doi.org/10.1175/MWR-D-15-0242.1).
- Birner, T., 2006: Fine-scale structure of the extratropical tropopause region. *J. Geophys. Res.*, **111**, D04104, doi:[10.1029/2005JD006301](https://doi.org/10.1029/2005JD006301).
- Brown, B. G., and G. S. Young, 2000: Verification of icing and turbulence forecasts: Why some verification statistics can't be computed using PIREPs. Preprints, *Ninth Conf. on Aviation, Range, and Aerospace Meteorology*, Orlando, FL, Amer. Meteor. Soc., 393–398.
- Brown, R., 1973: New indices to locate clear-air turbulence. *Meteor. Mag.*, **102**, 347–361.
- Calabrese, P. A., 1966: Forecasting mountain waves. U.S. Weather Bureau Tech. Memo. FCST-6, 12 pp.
- Cho, J. Y. N., R. E. Newell, B. E. Anderson, J. D. W. Barrick, and K. L. Thornhill, 2003: Characterizations of tropospheric turbulence and stability layers from aircraft observations. *J. Geophys. Res.*, **108**, 8784, doi:[10.1029/2002JD002820](https://doi.org/10.1029/2002JD002820).
- Colson, D., and H. A. Panofsky, 1965: An index of clear air turbulence. *Quart. J. Roy. Meteor. Soc.*, **91**, 507–513, doi:[10.1002/qj.49709139010](https://doi.org/10.1002/qj.49709139010).
- Cornman, L. B., 2016: Airborne in situ measurements of turbulence. *Aviation Turbulence: Processes, Detection, Prediction*, R. Sharman and T. Lane, Eds., Springer, 97–120, doi:[10.1007/978-3-319-23630-8_5](https://doi.org/10.1007/978-3-319-23630-8_5).
- , C. S. Morse, and G. Cunning, 1995: Real-time estimation of atmospheric turbulence severity from in-situ aircraft measurements. *J. Aircr.*, **32**, 171–177, doi:[10.2514/3.46697](https://doi.org/10.2514/3.46697).
- Doswell, C. A., III, R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576–585, doi:[10.1175/1520-0434\(1990\)005<0576:OSMOSI>2.0.CO;2](https://doi.org/10.1175/1520-0434(1990)005<0576:OSMOSI>2.0.CO;2).
- Doyle, J. D., and Coauthors, 2011: An intercomparison of T-REX mountain-wave simulations and implications for mesoscale predictability. *Mon. Wea. Rev.*, **139**, 2811–2831, doi:[10.1175/MWR-D-10-05042.1](https://doi.org/10.1175/MWR-D-10-05042.1).
- Dutton, M. J. O., 1980: Probability forecasts of clear-air turbulence based on numerical output. *Meteor. Mag.*, **109**, 293–310.
- Eckermann, S. D., J. Ma, and D. Broutman, 2004: The NRL Mountain Wave Forecast Model (MWFM). Preprints, *Symp. on the 50th Anniversary of Operational Numerical Weather Prediction*, College Park, MD, Amer. Meteor. Soc., P2.9. [Available online at <http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA465017>]
- , D. Broutman, J. Ma, and J. Lindeman, 2006: Fourier-ray modeling of short-wavelength trapped lee waves observed in infrared satellite imagery near Jan Mayen. *Mon. Wea. Rev.*, **134**, 2830–2848, doi:[10.1175/MWR3218.1](https://doi.org/10.1175/MWR3218.1).
- Ellrod, G. P., and D. I. Knapp, 1992: An objective clear-air turbulence forecasting technique: Verification and operational use. *Wea. Forecasting*, **7**, 150–165, doi:[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, doi:[10.1175/2009WAF2222290.1](https://doi.org/10.1175/2009WAF2222290.1).
- Endlich, R. M., 1964: The mesoscale structure of some regions of clear-air turbulence. *J. Appl. Meteor.*, **3**, 261–276, doi:[10.1175/1520-0450\(1964\)003<0261:TMSOSR>2.0.CO;2](https://doi.org/10.1175/1520-0450(1964)003<0261:TMSOSR>2.0.CO;2).
- Fahey, T. H., III, 1993: Northwest Airlines atmospheric hazards advisory and avoidance system. Preprints, *Fifth Int. Conf. on Aviation Weather Systems*, Vienna, VA, Amer. Meteor. Soc., 409–413.
- , M. Pfleiderer, and R. Sharman, 2002: Mountain wave activity and turbulence—Aviation forecasts and avoidance. Preprints, *10th Conf. on Aviation, Range, and Aerospace Meteorology*, Portland, OR, Amer. Meteor. Soc., 303–306.
- , E. N. Wilson, R. O'Loughlin, M. Thomas, and S. Klipfel, 2016: A history of weather reporting from aircraft and turbulence forecasting for commercial aviation. *Aviation Turbulence: Processes, Detection, Prediction*, R. Sharman and T. Lane, Eds., Springer, 31–58, doi:[10.1007/978-3-319-23630-8_2](https://doi.org/10.1007/978-3-319-23630-8_2).
- Frehlich, R., and R. Sharman, 2004a: Estimates of turbulence from numerical weather prediction model output with applications to turbulence diagnosis and data assimilation. *Mon. Wea. Rev.*, **132**, 2308–2324, doi:[10.1175/1520-0493\(2004\)132<2308:EOTFNW>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<2308:EOTFNW>2.0.CO;2).
- , and —, 2004b: Estimates of upper level turbulence based on second order structure functions derived from numerical weather prediction model output. Preprints, *11th Conf. on Aviation, Range, and Aerospace Meteorology*, Hyannis, MA, Amer.

- Meteor. Soc., 4.13. [Available online at <https://ams.confex.com/ams/pdfpapers/81831.pdf>.]
- , —, 2010: Climatology of velocity and temperature turbulence statistics determined from rawinsonde and ACARS/AMDAR data. *J. Appl. Meteor. Climatol.*, **49**, 1149–1169, doi:[10.1175/2010JAMC2196.1](https://doi.org/10.1175/2010JAMC2196.1).
- , Y. Meillier, M. L. Jensen, and B. Balsley, 2004: A statistical description of small-scale turbulence in the low-level nocturnal jet. *J. Atmos. Sci.*, **61**, 1079–1085, doi:[10.1175/1520-0469\(2004\)061<1079:ASDOST>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<1079:ASDOST>2.0.CO;2).
- , R. Sharman, F. Vandenbergh, W. Yu, Y. Liu, J. Kneivel, and G. Jumper, 2010: Estimates of C_n^2 from numerical weather prediction model output and comparison with thermosonde data. *J. Appl. Meteor. Climatol.*, **49**, 1742–1755, doi:[10.1175/2010JAMC2350.1](https://doi.org/10.1175/2010JAMC2350.1).
- Frisch, U., 1995: *Turbulence: The Legacy of A. N. Kolmogorov*. Cambridge University Press, 296 pp.
- Fritts, D. C., and M. J. Alexander, 2003: Gravity wave dynamics and effects in the middle atmosphere. *Rev. Geophys.*, **41**, 1003, doi:[10.1029/2001RG000106](https://doi.org/10.1029/2001RG000106).
- Gill, P. G., 2014: Objective verification of World Area Forecast Centre clear air turbulence forecasts. *Meteor. Appl.*, **21**, 3–11, doi:[10.1002/met.1288](https://doi.org/10.1002/met.1288).
- , 2016: Aviation turbulence forecast verification. *Aviation Turbulence: Processes, Detection, Prediction*, R. Sharman and T. Lane, Eds., Springer, 261–283.
- , and A. J. Stirling, 2013: Including convection in global turbulence forecasts. *Meteor. Appl.*, **20**, 107–114, doi:[10.1002/met.1315](https://doi.org/10.1002/met.1315).
- , and P. Buchanan, 2014: An ensemble based turbulence forecasting system. *Meteor. Appl.*, **21**, 12–19, doi:[10.1002/met.1373](https://doi.org/10.1002/met.1373).
- Guyon, I., and A. Elisseeff, 2003: An introduction of variable and feature selection. *J. Mach. Learn. Res.*, **3**, 1157–1182.
- Haltiner, G. J., and R. T. Williams, 1983: *Numerical Prediction and Dynamic Meteorology*. 2nd ed. John Wiley and Sons, 477 pp.
- Hopkins, R. H., 1977: Forecasting techniques of clear-air turbulence including that associated with mountain waves. WMO Tech. Note WMO/TN-155, 31 pp.
- ICAO, 2001: Meteorological service for international air navigation. Annex 3 to the Convention on International Civil Aviation, 14th Ed., ICAO Rep., 128 pp.
- Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons, 239 pp.
- Kaplan, M. L., and Coauthors, 2004: Characterizing the severe turbulence environments associated with commercial aviation accidents: A Real-Time Turbulence Model (RTTM) designed for the operational prediction of hazardous aviation turbulence environments. NASA Rep. NASA/CR-2004-213025, 54 pp. [Available online at <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20040110976.pdf>.]
- Kim, J.-H., and H.-Y. Chun, 2011: Statistics and possible sources of aviation turbulence over South Korea. *J. Appl. Meteor. Climatol.*, **50**, 311–324, doi:[10.1175/2010JAMC2492.1](https://doi.org/10.1175/2010JAMC2492.1).
- , —, 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, doi:[10.1175/JAMC-D-10-05017.1](https://doi.org/10.1175/JAMC-D-10-05017.1); Corrigendum, **50**, 2193, doi:[10.1175/JAMC-D-11-0188.1](https://doi.org/10.1175/JAMC-D-11-0188.1).
- , —, —, and S. B. Trier, 2014: The role of vertical shear on aviation turbulence within cirrus bands of a simulated western Pacific cyclone. *Mon. Wea. Rev.*, **142**, 2794–2813, doi:[10.1175/MWR-D-14-00008.1](https://doi.org/10.1175/MWR-D-14-00008.1).
- , W. N. Chan, B. Sridhar, and R. D. Sharman, 2015: Combined winds and turbulence prediction system for automated air-traffic management applications. *J. Appl. Meteor. Climatol.*, **54**, 766–784, doi:[10.1175/JAMC-D-14-0216.1](https://doi.org/10.1175/JAMC-D-14-0216.1).
- Kim, Y.-J., and A. Arakawa, 1995: Improvement of orographic gravity wave parameterization using a mesoscale gravity wave model. *J. Atmos. Sci.*, **52**, 1875–1902, doi:[10.1175/1520-0469\(1995\)052<1875:IOOGWP>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1875:IOOGWP>2.0.CO;2).
- , and J. D. Doyle, 2005: Extension of an orographic-drag parameterization scheme to incorporate orographic anisotropy and flow blocking. *Quart. J. Roy. Meteor. Soc.*, **131**, 1893–1921, doi:[10.1256/qj.04.160](https://doi.org/10.1256/qj.04.160).
- , S. D. Eckermann, and H.-Y. Chun, 2003: An overview of the past, present and future of gravity-wave drag parametrization for numerical climate and weather prediction models. *Atmos.–Ocean*, **41**, 65–98, doi:[10.3137/ao.410105](https://doi.org/10.3137/ao.410105).
- Knox, J. A., 1997: Possible mechanisms of clear-air turbulence in strongly anticyclonic flows. *Mon. Wea. Rev.*, **125**, 1251–1259, doi:[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, doi:[10.1175/2008JAS2477.1](https://doi.org/10.1175/2008JAS2477.1).
- , A. W. Black, J. A. Rackley, E. N. Wilson, J. S. Grant, S. P. Phelps, D. S. Nevius, and C. B. Dunn, 2016: Automated turbulence forecasting strategies. *Aviation Turbulence: Processes, Detection, Prediction*, R. Sharman and T. Lane, Eds., Springer, 243–260, doi:[10.1007/978-3-319-23630-8_12](https://doi.org/10.1007/978-3-319-23630-8_12).
- Krozel, J., V. Klimenko, and R. Sharman, 2011: Analysis of clear-air turbulence avoidance maneuvers. *Air Traffic Control Quart.*, **19**, 147–168. [Available online at <http://arc.aiaa.org/doi/pdf/10.2514/atcq.19.2.147>.]
- Laikthman, D. L., and Y. Z. Al'ter Zalik, 1966: Use of aerological data for determination of aircraft buffeting in the free atmosphere. *Izv. Akad. Nauk SSSR. Fiz. Atmos. Okeana*, **2**, 534–536.
- Lane, T. P., J. D. Doyle, R. Plougonven, M. A. Shapiro, and R. D. Sharman, 2004: Observations and numerical simulations of inertia–gravity waves and shearing instabilities in the vicinity of a jet stream. *J. Atmos. Sci.*, **61**, 2692–2706, doi:[10.1175/JAS3305.1](https://doi.org/10.1175/JAS3305.1).
- Lee, D. R., R. B. Stull, and W. S. Irvine, 1984: Clear air turbulence forecasting techniques. Air Force Global Weather Center Rep. AFGWC/TN-79-001, 76 pp. [Available online at www.dtic.mil/get-tr-doc/pdf?AD=ADA144854.]
- Lindborg, E., 1999: Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence? *J. Fluid Mech.*, **388**, 259–288, doi:[10.1017/S0022112099004851](https://doi.org/10.1017/S0022112099004851).
- MacCready, P. B., Jr., 1964: Standardization of gustiness values from aircraft. *J. Appl. Meteor.*, **3**, 439–449, doi:[10.1175/1520-0450\(1964\)003<0439:SOGVFA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1964)003<0439:SOGVFA>2.0.CO;2).
- 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.
- Marzban, C., 2004: The ROC curve and the area under it as performance measures. *Wea. Forecasting*, **19**, 1106–1114, doi:[10.1175/825.1](https://doi.org/10.1175/825.1).
- McCann, D. W., 2001: Gravity waves, unbalanced flow, and aircraft clear air turbulence. *Natl. Wea. Dig.*, **25** (1–2), 3–14. [Available

- online at <http://nwasfiles.nwas.org/digest/papers/2001/Vol25No12/Pg3-McCann.pdf>.]
- , J. A. Knox, and P. D. Williams, 2012: An improvement in clear-air turbulence forecasting based on spontaneous imbalance theory: The ULTURB algorithm. *Meteor. Appl.*, **19**, 71–78, doi:10.1002/met.260.
- Mogil, H. M., and R. L. Holle, 1972: Anomalous gradient winds: Existence and implications. *Mon. Wea. Rev.*, **100**, 709–716, doi:10.1175/1520-0493(1972)100<0709:AGWEAI>2.3.CO;2.
- Muñoz-Esparza, D., J. A. Sauer, R. R. Linn, and B. Kosović, 2016: Limitations of one-dimensional mesoscale PBL parameterizations in reproducing mountain-wave flows. *J. Atmos. Sci.*, **73**, 2603–2614, doi:10.1175/JAS-D-15-0304.1.
- Nakanishi, M., and H. Niino, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87**, 895–912, doi:10.2151/jmsj.87.895.
- Nastrom, G. D., and K. S. Gage, 1985: A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.*, **42**, 950–960, doi:10.1175/1520-0469(1985)042<0950:ACOAWS>2.0.CO;2.
- Nicholls, J. M., 1973: The airflow over mountains: Research 1958–1972. WMO Tech. Note WMO/TN-127, 73 pp.
- Palmer, T. N., G. J. Shutts, and R. Swinbank, 1986: Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parametrization. *Quart. J. Roy. Meteor. Soc.*, **112**, 1001–1039, doi:10.1002/qj.49711247406.
- Pan, L. L., W. J. Randel, B. L. Gary, M. J. Mahoney, and E. J. Hintsa, 2004: Definitions and sharpness of the extratropical tropopause: A trace gas perspective. *J. Geophys. Res.*, **109**, D23103, doi:10.1029/2004JD004982.
- Pearson, J. M., and R. D. Sharman, 2017: Prediction of energy dissipation rates for aviation turbulence. Part II: Nowcasting convective and nonconvective turbulence. *J. Appl. Meteor. Climatol.*, **56**, 339–351, doi:10.1175/JAMC-D-16-0312.1.
- Pepe, M. S., and M. L. Thompson, 2000: Combining diagnostic test results to increase accuracy. *Biostatistics*, **1**, 123–140, doi:10.1093/biostatistics/1.2.123.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1986: *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 848 pp.
- Reap, R. M., 1996: Probability forecasts of clear-air-turbulence for the contiguous US. National Weather Service Office of Meteorology Tech. Procedures Bull. 430, 15 pp. [Available online at http://www.nws.noaa.gov/mdl/pubs/Documents/TechProcBulls/TPB_430.pdf.]
- Roach, W. T., 1970: On the influence of synoptic development on the production of high level turbulence. *Quart. J. Roy. Meteor. Soc.*, **96**, 413–429, doi:10.1002/qj.49709640906.
- Schumann, U., 2012: A contrail cirrus prediction model. *Geosci. Model Dev.*, **5**, 543–580, doi:10.5194/gmd-5-543-2012.
- Sharman, R., and T. Lane, Eds., 2016: *Aviation Turbulence: Processes, Detection, Prediction*. Springer, 523 pp., doi:10.1007/978-3-319-23630-8.
- , C. Tebaldi, G. Wiener, and J. Wolff, 2006: An integrated approach to mid- and upper-level turbulence forecasting. *Wea. Forecasting*, **21**, 268–287, doi: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, doi:10.1029/2012GL051996.
- , L. B. Cornman, G. Meymaris, J. Pearson, and T. Farrar, 2014: Description and derived climatologies of automated in situ eddy-dissipation-rate reports of atmospheric turbulence. *J. Appl. Meteor. Climatol.*, **53**, 1416–1432, doi:10.1175/JAMC-D-13-0329.1.
- Shutts, G., 1997: Operational lee wave forecasting. *Meteor. Appl.*, **4**, 23–35, doi:10.1017/S1350482797000340.
- Stone, P. H., 1966: On non-geostrophic baroclinic stability. *J. Atmos. Sci.*, **23**, 390–400, doi:10.1175/1520-0469(1966)023<0390:ONGBS>2.0.CO;2.
- Trier, S. B., and R. Sharman, 2016: Mechanisms influencing cirrus banding and aviation turbulence near a convectively enhanced upper-level jet stream. *Mon. Wea. Rev.*, **144**, 3003–3027, doi:10.1175/MWR-D-16-0094.1.
- , —, and T. P. Lane, 2012: Influences of moist convection on a cold-season outbreak of clear-air turbulence (CAT). *Mon. Wea. Rev.*, **140**, 2477–2496, doi:10.1175/MWR-D-11-00353.1.
- Turner, J., 1999: Development of a mountain wave turbulence prediction scheme for civil aviation. Met Office Tech. Rep. 265, 34 pp.
- Vosper, S., 2003: Development and testing of a high resolution mountain-wave forecasting system. *Meteor. Appl.*, **10**, 75–86, doi:10.1017/S1350482703005085.
- Williams, J. K., 2014: Using random forests to diagnose aviation turbulence. *Mach. Learn.*, **95**, 51–70, doi:10.1007/s10994-013-5346-7.
- , and G. Meymaris, 2016: Remote turbulence detection using ground-based Doppler weather radar. *Aviation Turbulence: Processes, Detection, Prediction*, R. Sharman and T. Lane, Eds., Springer, 149–177, doi:10.1007/978-3-319-23630-8_7.
- Wurtele, M. G., R. Sharman, and A. Datta, 1996: Atmospheric lee waves. *Annu. Rev. Fluid Mech.*, **28**, 429–476, doi:10.1146/annurev.fl.28.010196.002241.
- Zovko-Rajak, D., and T. P. Lane, 2014: The generation of near-cloud turbulence in idealized simulations. *J. Atmos. Sci.*, **71**, 2430–2451, doi:10.1175/JAS-D-13-0346.1.