An Improved Algorithm for Low-Level Turbulence Forecasting

DOMINGO MUÑOZ-ESPARZA AND ROBERT SHARMAN

National Center for Atmospheric Research,^a Boulder, Colorado

(Manuscript received 3 December 2017, in final form 16 February 2018)

ABSTRACT

A low-level turbulence (LLT) forecasting algorithm is proposed and implemented within the Graphical Turbulence Guidance (GTG) turbulence forecasting system. The LLT algorithm provides predictions of energy dissipation rate (EDR; turbulence dissipation to the one-third power), which is the standard turbulence metric used by the aviation community. The algorithm is based upon the use of distinct log-Weibull and lognormal probability distributions in a statistical remapping technique to represent accurately the behavior of turbulence in the atmospheric boundary layer for daytime and nighttime conditions, respectively, thus accounting for atmospheric stability. A 1-yr-long GTG LLT calibration was performed using the High-Resolution Rapid Refresh operational model, and optimum GTG ensembles of turbulence indices for clear-air and mountain-wave turbulence that minimize the mean absolute percentage error (MAPE) were determined. Evaluation of the proposed algorithm with in situ EDR data from the Boulder Atmospheric Observatory tower covering a range of altitudes up to 300 m above the surface demonstrates a reduction in the error by a factor of approximately 2.0 (MAPE = 55%) relative to the current operational GTG system (version 3). In addition, the probability of detection of typical small and large EDR values at low levels is increased by approximately 15%–20%. The improved LLT algorithm is expected to benefit several nonconventional turbulence-prediction sectors such as unmanned aerial systems and wind energy.

1. Introduction

Turbulence forecasts are used by pilots, dispatchers, and air traffic controllers to strategically avoid regions of strong turbulence that can lead to injuries, aircraft fatigue, and damage (Sharman and Lane 2016). Current turbulence forecast methods use numerical weather prediction (NWP) model output to diagnose different sources of turbulence, including clear-air turbulence (CAT), mountain-wave turbulence (MWT), and convectively induced turbulence at mid- and upper levels, and low-level turbulence (LLT) owing to turbulence events within the atmospheric boundary layer. These turbulence forecasting techniques typically relate the large-scale atmospheric turbulence production mechanisms that are explicitly

resolved with horizontal grid spacings of NWP models (~ 10 km) to aircraft-scale turbulence (~ 10 – 100 m), through the assumption of a downscale energy-cascade process (e.g., Lindborg 1999). An example of turbulence forecasting algorithm is the Graphical Turbulence Guidance (GTG) system, originally devised by Sharman et al. (2006), that is run operationally for public use by NOAA's National Weather Service (<https://www.aviationweather.gov/turbulence/gtg>). GTG uses an ensemble average that is based on an optimum combination of turbulence diagnostics that has been recently extended by (Sharman and Pearson 2017) to 1) provide forecasts at all flight altitudes from surface to flight level (FL) 500 (50 000 ft; 1 ft = 30.48 cm), 2) explicitly provide forecasts of MWT, and 3) output energy dissipation rate (EDR; $=\epsilon^{1/3}$, where ϵ is the turbulence dissipation rate) instead of a likelihood metric (i.e., categorical levels such as light, moderate, or severe turbulence). In addition, GTG has been extended to perform EDR nowcasts using a short-term forecast nudged to observations (Pearson and Sharman 2017). This approach has shown to be particularly useful in identifying areas of turbulence associated with convection.

The original motivation for the development of GTG was to create an operational system that could be used

Denotes content that is immediately available upon publication as open access.

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

Corresponding author: Dr. Domingo Muñoz-Esparza, domingom@ucar.edu

as a guide for making turbulence-avoidance decisions for the aviation community. Therefore, emphasis was placed on forecasting turbulence primarily at upper levels (20 000–45 000 ft: FL 200–450) and midlevels (10 000–20 000 ft), for turbulence levels relevant to commercial aircraft. Within these altitude ranges, climatological distributions derived from in situ aircraft and pilot reports (PIREPs) support the notion that EDR follows a lognormal distribution (e.g., [Nastrom and Gage 1985](#); [Cho et al. 2003](#); [Frehlich and Sharman 2004](#); [Sharman et al. 2014](#)). In light of these findings, [Sharman and Pearson \(2017\)](#) proposed a remapping method that is based on the lognormal nature of upper-tropospheric and lower-stratospheric (UTLS) EDR (the technique is outlined in [section 2](#)). Turbulence generation mechanisms within the atmospheric boundary layer (ABL) differ from those at the UTLS region, however, mainly because of the proximity to Earth's surface and the role of the surface energy budget. A recent study by [Muñoz-Esparza et al. \(2018\)](#) has shown that, in the first few hundred meters above the surface, EDR is strongly driven by the diurnal evolution of the ABL. During convective daytime ABL conditions, EDR bears a log-Weibull distribution, whereas nighttime stably stratified ABLs follow a lognormal distribution. The similarity between LLT during stable conditions and UTLS turbulence in terms of statistical behavior is attributed to upper levels being typically characterized by quasi two-dimensional stratified turbulence (e.g., [Lindborg 1999](#)). Therefore, LLT predictions may necessitate a special treatment that differs from the mid- and upper-level turbulence forecast methods.

GTG has focused up to now on mid- and upper-level turbulence aviation forecasts, but there are other existing and emerging areas that can specifically benefit from improved strategic LLT forecasts. From an aviation standpoint, the use of small unmanned aerial systems (UAS), or drones, is growing at a fast pace. A clear example of the proliferation of such technology is the initiative of NASA to develop an air traffic management system for UASs (<https://utm.arc.nasa.gov/>; [Kopardekar et al. 2016](#)), with a main focus of enabling safe yet efficient low-altitude operations for UASs [in the lowest 200–500 ft (~ 60 – 150 m)]. In particular, such systems are more than an order of magnitude smaller in size and weight than typical commercial aircraft and therefore are expected to be sensitive to smaller-magnitude turbulence encounters that do not affect commercial aircraft operations at upper levels. Another aviation application of relevance is wake vortices (e.g., [Holzapfel et al. 2016](#)). Vortices that form in the wake of an aircraft affect the latency of takeoff/landing maneuvers, as well as its design and execution. Interaction with downstream

aircraft can result in wake-vortex encounters that pose difficulties in aircraft maneuvers and can even lead to aircraft accidents. The extent of these wake vortices depends on the background atmospheric LLT, with smaller ABL EDRs resulting in longer-lived wake vortices. Also, atmospheric LLT has significant impacts on wind turbine wake characteristics and thus affects power production and wind-plant control. In particular, recent field studies have found that the nature of the turbine wake strongly depends on background EDR of the incoming ABL ([Smalikho et al. 2013](#); [Lundquist and Bariteau 2015](#)).

All these examples evidence the relevance and potential benefits of operational LLT forecasting algorithms. Herein, we propose a new method that extends the capabilities of GTG to improve the prediction of EDR in the ABL. The remainder of the paper is structured as follows. The new algorithm to predict low-level EDR that accounts for turbulence variability throughout the diurnal cycle of the ABL is introduced in [section 2](#). The particular turbulence diagnostics utilized, as well as the calibration of the algorithm and determination of the optimum ensemble forecast, are discussed in [section 3](#). Validation of the proposed LLT forecast method and its comparison with the predecessor GTG are presented in [section 4](#). [Section 5](#) is devoted to the conclusions.

2. A new approach to account for diurnal variability of turbulence in the ABL

The current GTG system uses a statistical remapping technique ([Sharman and Pearson 2017](#)). The method transforms any turbulence index or diagnostic of arbitrary units D_i derived from an NWP model into an EDR value D_i^* ($\text{m}^{2/3}\text{s}^{-1}$), applying a linear remapping approach in logarithmic space:

$$\ln D_i^* = a + b \ln D_i. \quad (1)$$

The coefficients a and b in Eq. (1) are calculated on the basis of the expectation operator $\langle \rangle$ and the standard deviation operator $\text{SD}()$ of the probability distribution function (PDF) of the observations and the quantity to be remapped, using the following expressions:

$$b = \frac{\text{SD}(\ln \varepsilon^{1/3})}{\text{SD}(\ln D_i)} = \frac{C_2}{\text{SD}(\ln D_i)} \quad \text{and} \quad (2)$$

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

where $\varepsilon^{1/3}$ represents the observational “climatology” of reference.

[Muñoz-Esparza et al. \(2018\)](#) have recently found using data from the Experimental Planetary Boundary

Layer Instrumentation Assessment (XPIA) field campaign (Lundquist et al. 2017) that probability distributions in the ABL exhibit distinct patterns during daytime and nighttime conditions. During daytime convective ABLs (i.e., CBLs), the probability distribution of EDR follows a log-Weibull distribution, characterized by

$$\text{PDF}(\ln x) = \frac{k}{\lambda} \left(\frac{\ln x}{\lambda} \right)^{k-1} \exp \left[- \left(\frac{\ln x}{\lambda} \right)^k \right], \quad (4)$$

where k and λ are the shape and scale parameters of the distribution, respectively. For a log-Weibull distribution, the mean and standard deviation are given by $\langle \ln x \rangle = \lambda \Gamma(1 + k^{-1})$ and $\text{SD}(\ln x) = \lambda \{ \Gamma(1 + 2k^{-1}) - [\Gamma(1 + k^{-1})]^2 \}^{1/2}$, respectively, where Γ is the gamma function. In contrast, nighttime stable ABL (i.e., SBL) conditions bear a lognormal distribution as at upper levels, defined as

$$\text{PDF}(\ln x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp \left[- \frac{(\ln x - \langle \ln x \rangle)^2}{2\sigma^2} \right], \quad (5)$$

with the expectation and standard deviation of the distribution given by $\langle \ln x \rangle = \exp(\mu + 0.5\sigma^2)$ and $\text{SD}(\ln x) = \exp(\mu + 0.5\sigma^2)[\exp(\sigma^2) - 1]^{1/2}$, respectively. The EDR database generated by Muñoz-Esparza et al. (2018) encompasses levels from the surface to $z = 300$ m, covering 3 months (March–May 2015) of observations at the Boulder Atmospheric Observatory (BAO) tower near Boulder, Colorado (40°03'N, 105°00'W). The tower was instrumented with 20-Hz sonic anemometers at seven heights, resulting in a database (“climatology”) of ~ 1.5 million EDR records.

Figure 1a shows the mean and standard deviation of the probability distributions from Muñoz-Esparza et al. (2018) as a function of height above the surface and sorted by atmospheric stability. Solid lines represent the best fit to power and linear functions for the mean and standard deviation of the distributions, respectively. These were used to derive the mean parameters of the observational climatologies for the vertical grid points that are within the first 300 m above the surface in the High-Resolution Rapid Refresh (HRRR; Smith et al. 2008; <http://ruc.noaa.gov/hrrr/>) operational forecasting system that runs over the contiguous United States (CONUS) and is based on the Weather Research and Forecasting (WRF) Model (Skamarock and Klemp 2008). For the HRRR model, this includes five vertical model levels approximately located at $z = 10, 37.5, 85, 165$, and 280 m. Note that these are very similar to the Rapid Refresh (RAP; Benjamin et al. 2016) forecast

FIG. 1. (a) Mean and standard deviation of the EDR PDFs at the BAO tower during the XPIA campaign from Muñoz-Esparza et al. (2018) (dots). Solid lines correspond to the best fit of power and linear functions for the mean and standard deviation of the distributions, respectively. (b) Resulting reference PDFs for convective and stable ABL conditions as based on height-averaged distribution parameters for the following heights above surface level: $z = 10, 37.5, 85, 165$, and 280 m. The height-averaged values for the log-Weibull daytime distribution are $C_1 = -2.1980 \text{ m}^{2/3} \text{ s}^{-1}$ and $C_2 = 0.6564 \text{ m}^{2/3} \text{ s}^{-1}$ ($k = -5.3$; $\lambda = -1.9$), and nighttime lognormal distribution values are $C_1 = -2.7017 \text{ m}^{2/3} \text{ s}^{-1}$ and $C_2 = 0.7641 \text{ m}^{2/3} \text{ s}^{-1}$.

product, with a slight increase of the vertical spacing with respect to the HRRR of about 1.3%; therefore the same parameters can be used for both NWP forecast models as input for GTG. The height-averaged values for the log-Weibull daytime distribution are $C_1 = -2.1980 \text{ m}^{2/3} \text{ s}^{-1}$ and $C_2 = 0.6564 \text{ m}^{2/3} \text{ s}^{-1}$, whereas nighttime lognormal distribution values are $C_1 = -2.7017 \text{ m}^{2/3} \text{ s}^{-1}$ and $C_2 = 0.7641 \text{ m}^{2/3} \text{ s}^{-1}$. The resulting PDFs for CBL and SBL conditions as based on the previous height-averaged parameters are presented in Fig. 1b. These results clearly indicate a predominance of large EDR values during daytime conditions, whereas nighttime LLT tends to be weaker. We explored the use of height-dependent coefficients for the PDFs, but only minor differences were

found, likely because of other sources of error being more dominant. Therefore, height-averaged coefficients are used for simplicity.

The main changes with respect to the previous GTG algorithm for LLT are the use of two different PDFs for daytime and nighttime conditions (log-Weibull and lognormal, respectively) and the fact that the reference PDFs are derived from ABL measurements using sonic anemometers instead of in situ aircraft and PIREPs. The advantages of this new reference database are discussed in section 4. In addition, the LLT algorithm uses the total surface heat flux H (W m^{-2}), which is the sum of the sensible and latent heat fluxes, to discriminate between convective ($H \geq 0$) and stable ($H < 0$) conditions and therefore uses the appropriate PDF for the statistical remapping [Eq. (1)]. Moreover, two different suites of indices are used for daytime and nighttime conditions. This approach is beneficial since turbulence mechanisms, and therefore GTG indices, depend on ABL stability, which results in an improvement of forecast skill at low levels in comparison with the state-of-the-art GTG algorithm from Sharman and Pearson (2017).

3. Turbulence diagnostics, calibration, and optimum ensemble forecast

As already mentioned, GTG uses operational NWP model output as a basis for computing all of the turbulence indices, typically involving calculation of spatial gradients. The current operational version of GTG utilizes the WRF-based RAP model, which has a horizontal grid spacing of ~ 13 km and covers the CONUS. For global applications, GTG uses 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. The WRF-based HRRR product is also available, which has a finer horizontal grid spacing of 3 km. For the next operational release of GTG we will make a transition to the HRRR model, since we expect to improve our EDR forecasts by better representing terrain effects and therefore flow features that lead to turbulence production. Therefore, for the development of the new LLT algorithm we use the HRRR model as the input operational NWP model for GTG.

The HRRR model output is normally available every hour from the current time out to an 18-h forecast horizon. In a typical case, the accuracy of the forecast progressively degrades as the forecast lead time increases. Mesoscale NWP models, however, require a spinup time to generate the energy content in the mesoscale portion of the energy spectrum that is missing in the initial conditions, mostly containing energy at the

synoptic scales. In GTG, we want to have such mesoscale variability developed, which will subsequently be related to turbulence at smaller scales that are not resolved by the model. To quantify the spinup time of the HRRR product, energy spectra for the horizontal wind speed were computed:

$$F_U(k) = \left| \sum_{n=0}^{N-1} U_n \exp(-ikn\Delta) \right|^2, \quad (6)$$

where k is the wavenumber, n is the index in a given spatial direction, and Δ is the corresponding grid spacing. Energy spectra of horizontal wind at 2 km above the surface are presented in Fig. 2. These results are valid at the same time, progressively decreasing the initialization time and increasing the forecast time in the same time increment, and correspond to 0, 1, 2, 3, 5, 7, and 9 h forecasts. The two cases correspond to a daytime instance (2000 UTC 9 March 2015; Fig. 2a) and a nighttime instance (1000 UTC 10 March 2015; Fig. 2b). In both cases, the initial-condition state displays a considerable energy underprediction at the mesoscales, as expected, which progressively grows with forecast time. During daytime conditions, the model reaches equilibrium (i.e., mesoscale energy content becomes fully developed) in ~ 3 h after its initialization. For the nighttime case, the development process appears to be slightly slower, requiring ~ 5 h for the energy spectrum to stabilize. We hypothesize that the presence of a convective source at the mesoscales accelerates the model spinup phase, which during nighttime is essentially dominated by an energy-cascade process and thus necessitates larger times to propagate across the entire resolvable spectrum of spatial scales. We only presented two cases to illustrate the typical behavior, but consistent conclusions were found for other days and times (not shown). On the basis of these results, forecasts with 5-h lead time were used to calibrate GTG LLT indices and for the statistical analysis presented in section 4. This time is in fact very similar to the 6-h forecast used by Sharman and Pearson (2017) to generate climatological distributions of turbulence indices. Note also that the model displays an energy dependence on the wavelength of $\sim k^{-5/3}$ at the mesoscales, consistent with both observations and modeling studies (e.g., Lindborg 1999; Skamarock 2004; Skamarock et al. 2014).

Sharman and Pearson (2017) provide a list of all of the turbulence indices in GTG. In addition to these, two indices that are specific to ABL turbulence were incorporated. One is an EDR estimate that is derived from a turbulent kinetic energy (TKE; q) budget model that is based on large-eddy simulations for convective ABLs ("MSEDR"; Moeng and Sullivan 1994):

FIG. 2. Horizontal wind energy spectra at a height of ~ 2 km above the surface for (a) 2000 UTC 9 Mar and (b) 1000 UTC 10 Mar 2015. Results are valid at the same time, but with different lead times, and correspond to 0, 1, 2, 3, 5, 7, and 9 h forecasts. Spectra are calculated over the entire HRRR grid over the meridional direction and are averaged in zonal direction.

$$\begin{aligned}\varepsilon &= -0.4w_*^3/z_i - u_*^3(1 - z/z_i)\phi_m/(\kappa z), \\ \phi_m &= (1 - 16\zeta)^{-1/4},\end{aligned}\quad (7)$$

where u_* is the friction velocity, w_* is the convective velocity scale, z_i is the ABL depth, ϕ_m is the surface layer stability function, $\zeta = z/L$, $L = -u_*^3\theta\rho c_p/(\kappa gH)$ is the Obukhov length, θ is the potential temperature, ρ is the density, c_p is the specific heat of air at constant pressure, and κ is the von Kármán constant. The other additional index is from the dissipation-term parameterization in the planetary boundary layer (PBL) scheme that is based on Kolmogorov's hypothesis ("PBLEDR"):

$$\varepsilon = q^{3/2}/(b_1\ell), \quad (8)$$

where ℓ is the turbulent master length scale and b_1 is a model constant ($=24.0$). Muñoz-Esparza et al. (2018) have shown that WRF simulations using the same PBL scheme that the HRRR model uses—the Mellor–Yamada–Nakanishi–Niino (MYNN) scheme (Nakanishi 2001; Nakanishi and Niino 2006)—can produce estimates of EDR with a bias reduction of $\sim 75\%$ after applying the statistical remapping technique described in section 2. The HRRR output unfortunately only includes TKE, and therefore ℓ has to be calculated from the output. The turbulent length scale is computed as follows:

$$\begin{aligned}\ell &= \ell_1(1 - \alpha) + 0.4\alpha\ell_2, \\ \alpha &= 0.5 \tanh[(z - z_{i2} - h_1)/h_2] + 0.5,\end{aligned}\quad (9)$$

where $z_{i2} = \max(z_i, 300)$, $h_1 = \min(z_{i2}, 750)$, and $h_2 = 0.5h_1$. In Eq. (9), ℓ_2 is the turbulent length scale from Bougeault and Lacarrere (1989) and ℓ_1 corresponds to the

length scale from Nakanishi (2001), which is the harmonic mean of the surface, turbulence, and buoyancy scales:

$$\frac{1}{\ell_1} = \frac{1}{\ell_s} + \frac{1}{\ell_T} + \frac{1}{\ell_B}, \quad (10)$$

where

$$\begin{aligned}\ell_s &= \kappa z(1 - 20\zeta)^{0.2}, \\ \ell_T &= 0.23 \frac{\int_0^{z_i} qz dz}{\int_0^{z_i} q dz}, \quad \text{and} \\ \ell_B &= \ell_\infty,\end{aligned}$$

with $\ell_\infty = 1 \times 10^8$ m. The expression for ℓ_1 from Eq. (10) is valid for convective conditions. Although the master length scale also has a definition for stable ABLs, we found that Eq. (8) performed poorly under such stability, likely because of the coarse vertical grid spacing in the HRRR model, and therefore it is not included here.

To derive PDFs for the 77 indices available in GTG, we generated predictions for 1 yr (May 2016–April 2017) using the HRRR model for 5-h forecasts initialized every hour (24 times per day). We intentionally used HRRR data from a different period for two reasons: 1) to calibrate GTG with the most recent HRRR data available, since the HRRR model continuously undergoes tuning and modifications, and 2) to not include the calibration data as part of the training. A 60×60 km² area surrounding the BAO tower location was considered, including heights < 3 km above the surface. With that, and taking into account the availability of archived HRRR

FIG. 3. One-year probability distributions of 5-h forecasts with the HRRR model for the indices (a) MSEDR (Moeng and Sullivan 1994) during daytime conditions and (b) Ellrod2 (Ellrod and Knapp 1992) during nighttime conditions. Curves represent the best fit to log-Weibull [Eq. (4)] and lognormal [Eq. (5)] distributions for (a) and (b), respectively. Also shown are (c),(d) the same PDFs as in (a) and (b) but sorted by season. The total number of samples is $n = 45\,619\,200$.

output, a database of 45 619 200 samples was generated. Figures 3a and 3b show PDFs for two of the indices, MSEDR and “Ellrod2” (Ellrod and Knapp 1992), filtered for daytime and nighttime conditions, respectively, to illustrate the general behavior of the probability distributions. The PDFs are in good agreement with the observed log-Weibull and lognormal structure of ABL turbulence. Fitting of the PDFs to the analytical expressions is needed to determine the mean and standard deviation of the distributions. For that purpose, a least squares error minimization technique was used, considering bins with probability of occurrence larger than 0.1% (dark-blue dots). Nevertheless, this is a relatively minor correction, since the largest PDF values dominate the error minimization.

A limitation of the observational dataset used herein is the time span of the observations (March–May 2015). To investigate the likelihood of the sonic-anemometer-

derived PDFs to account properly for intra-annual variability, we use the 1-yr-long GTG predictions as a proxy and filter the indices by season to analyze the variability. For daytime conditions (Fig. 3c), there is good overall agreement, with a slight shift to larger EDR values during the spring and a higher small EDR tail for the distribution during winter conditions. We attribute these to weaker convective ABLs developing during the winter, as opposed to the spring and summer months. Nighttime turbulence presents even less intra-annual variability, with slightly weaker turbulence during summertime. These relatively small differences affirm the statistical robustness of the observations used as reference for the expected EDR distributions. Also, we emphasize that springtime downslope turbulence events are prone to occur at the BAO site for the period during which observations were available. In contrast to the midlevel and upper level, the ABL is continuously turbulent, and

therefore PDFs are dominated by turbulence events mainly driven by the stability of the ABL; these are well captured with a 3-month-long dataset.

The final GTG for CAT and MWT is computed as the ensemble average of the optimum indices:

$$\text{GTG} = \frac{1}{m} \sum_{i=1}^m w_i D_i^*. \quad (11)$$

In Eq. (11), m is the number of indices used by the specific combination; w_i is a weighting coefficient that can be adjusted dynamically or can be statically prescribed on the basis of performance for a long-term retrospective period (Sharman et al. 2006). Here, we give the same weight to all of the indices, $w_i = 1.0$, since the use of equal weights has been shown to provide sufficient accuracy in the latest GTG operational release. To determine the optimum suite of turbulence indices/diagnostics, different criteria can be employed. Sharman and Pearson (2017) derived the combination of GTG indices that maximizes the area under the curve (AUC) in “relative operating characteristic” curves for a threshold of $\text{EDR} = 0.22 \text{ m}^{2/3} \text{ s}^{-1}$, corresponding to the median 1-min peak EDR for moderate turbulence reports (Sharman et al. 2014). This criterion has been also found to be optimum in studies in which the main emphasis was on forecasting events above or below a given threshold (i.e., categorical classification) but was not necessarily focused on an accurate forecast along a wide range of values (e.g., Marzban 2004; Sharman et al. 2006; Gill 2014).

As mentioned in the introduction, an accurate forecast of both large and small EDR values is desirable for LLT applications. Thus, we optimized the GTG combination to minimize the mean absolute percentage error (MAPE), defined as

$$\text{MAPE}_i = \frac{100}{n} \sum_{i=1}^n \frac{|\text{EDR}_{\text{obs}} - D_i^*|}{\text{EDR}_{\text{obs}}}, \quad (12)$$

where n is the number of observation–forecast pairs available, D_i^* is the remapped index into EDR space, and EDR_{obs} is the corresponding observation. Such a statistical measure is desirable since the range of EDR values spans a few orders of magnitude, and therefore larger EDR values will dominate the optimization processes when using a nonrelative error metric. Moreover, CAT and MWT combinations are determined separately for daytime and nighttime conditions. This is done because of the differences in turbulence mechanisms for CBLs and SBLs (i.e., buoyant vs shear production), which in turn results in different statistical behavior and therefore different performances depending on atmospheric stability. Following Sharman and Pearson (2017), the optimization

process is performed by employing a forward selection of turbulence diagnostics starting with the one having the smallest MAPE and adding diagnostics that decrease the MAPE of the ensemble combination [Eq. (11)] until no further decrease of MAPE is achieved. A list of the subset of optimum indices used for the GTG LLT ensemble-average CAT and MWT combinations is presented in Table 1.

It is worth mentioning that, for daytime conditions, the term $b \ln D_i$ is replaced by $b(\ln D_i - \alpha)$ in Eq. (1). This is imposed to ensure that $\ln D_i - \alpha > 1$, since otherwise the gamma function in the log-Weibull distribution is not always a real number. This transformation does not affect the results, and it is accounted for in the a and b coefficients presented in Table 2, for which $\alpha = 25$ was used. Also, and although not explicitly indicated in the index definitions, the CAT indices for SBL conditions were divided by the Richardson number, as proposed in Sharman and Pearson (2017) to improve the statistical performance of the indices.

The suite of indices used in the optimum CAT and MTW combinations is presented in Table 2, which includes the individual MAPEs and the improvement of the intermediate GTG combinations as new indices are incorporated. Optimization is performed using EDR observations/forecasts at the BAO tower for $z = 5, 50, 100, 150$, and 300 m during April 2015 ($n = 3112$). Despite the difference in altitude from upper to lower levels, the stable ABL shares many commonalities with upper-level turbulence generation mechanisms. Both are stratified, the turbulence is quasi two-dimensional, and there is typically strong vertical shear. Therefore, indices that account for these mechanisms are potentially good candidates to diagnose low-level turbulence in stable conditions.

The use of an ensemble of indices improves the performance from any single individual index. For daytime conditions, PBLEDR is the best-performing CAT index, with $\text{MAPE} = 40.246\%$, further confirming that the simplified turbulent length scale that is used works appropriately. The MWT CBL combination has $\sim 20\%$ larger error than that for the CAT CBL. We attribute this lower performance to the fact that mountain waves are dampened in the presence of deeper boundary layers and larger turbulence levels (e.g., Smith and Skillingstad 2011; Sauer et al. 2016), as typically occur during daytime conditions. In contrast, for nighttime conditions CAT and MWT ensembles exhibit essentially the same performance, with a MAPE that is $\sim 15\%$ larger than in the case of the CAT CBL ensemble. We speculate that this is in part attributed to the coarse vertical grid resolution of the HRRR model and to difficulties of mesoscale models in stable ABLs either to resolve phenomena such as

TABLE 1. Definitions of low-level turbulence indices for CAT and MWT. Here NCSU1 is the North Carolina State University index, NGM indicates derivation from the Nested Grid Model, CTSQ stands for the thermal structure constant C_T^2 , WSQ is vertical velocity squared, and d_s is the near-source diagnostic.

Index	Description	Unit	Reference
DIV	Absolute value of the horizontal divergence	s^{-1}	
Ellrod2	Turbulence index 2 (TI2; includes divergence)	s^{-2}	Ellrod and Knapp (1992)
SCHGW	$\varepsilon^{1/3}$ derived from variance of vertical velocity	$m^{2/3} s^{-1}$	Schumann (2012)
NCSU1	Advection $\times$ absolute value of the vorticity gradient	$s^{-3} s^{-1}$	Kaplan et al. (2006)
NGM1	Wind speed $\times$ absolute value of the deformation	$m s^{-2}$	Reap (1996)
−NGM2	$-1 \times$ absolute value of the vertical temperature gradient $\times$ absolute value of the deformation	$K m^{-1} s^{-1}$	Reap (1996)
F3D	3D frontogenesis function on constant z surfaces	$K m s^{-1}$	Sharman et al. (2006)
SIGWAVG	Variance of w estimated from second-order structure functions	$m^{-2} s^{-2}$	Frehlich and Sharman (2004)
CTSQ	θ structure constant estimated from second-order structure functions	$K^2 m^{-2/3}$	Frehlich and Sharman (2010)
PBLED	$\varepsilon^{1/3}$ derived from TKE transport equation	$m^{2/3} s^{-1}$	Nakanishi (2001)
MSEDR	$\varepsilon^{1/3}$ derived from TKE budget model for CBL conditions	$m^{2/3} s^{-1}$	Moeng and Sullivan (1994)
MWT1	$d_s \times$ WSQ	$m^3 s^{-3}$	Sharman and Pearson (2017)
MWT3	$d_s \times$ F3D	$K m^2 s^{-2}$	Sharman and Pearson (2017)
MWT5	$d_s \times$ DIV	$m s^{-2}$	Sharman and Pearson (2017)
MWT7	$d_s \times$ SIGWAVG	$m^3 s^{-3}$	Sharman and Pearson (2017)
MWT11	$d_s \times$ SCHGW	$m^{5/3} s^{-2}$	Sharman and Pearson (2017)

low-level jets or to parameterize subgrid-scale processes such as turbulence intermittency, gravity waves, and Kelvin–Helmholtz instabilities (e.g., Holtslag et al. 2013; Muñoz-Esparza et al. 2017).

4. Validation and performance assessment

In this section, we validate the new LLT algorithm implemented in GTG using part of the XPJA campaign observations. The training of the algorithm was carried out with XPJA data from April of 2015, and the rest of the campaign was left for validation and quantification of forecast skill. Because of outages in the archived HRRR database, model output was unfortunately unavailable for May of 2015, therefore reducing the data for validation to the period 9–30 March 2015 ($n = 2462$). The RAP-based GTG ensemble results from Sharman and Pearson (2017) are also included to judge the relative performance of the new LLT algorithm. Both the HRRR and the RAP forecasts are based on the WRF Model, which uses a terrain-following pressure-based vertical coordinate. Such a choice results in a vertical coordinate that slightly changes position in time; for the specific period and heights used herein that effect has standard deviations of less than 2% of the mean value (height variability is indicated at the top of each panel in the following figures). To avoid potential errors in performing interpolation to match exactly the observation heights, the closest vertical grid point available from GTG was used.

Figures 4 and 5 show time evolutions of the GTG ensemble during the evaluation period for the CAT and

MWT indices, respectively. Visual inspection reveals several aspects related to the improved performance of the new LLT algorithm (GTG LLT HRRR, also referred to herein as GTG LLT). At all heights, GTG LLT exhibits a better diurnal evolution pattern, capturing the maximum EDR peaks occurring in the early evening and the nighttime minima. The use of different PDFs for daytime and nighttime conditions that are also distinct in structure, together with two different sets of indices depending upon ABL stability, results in a realistic capturing of the diurnal evolution of EDR in the ABL. In contrast, with the original GTG algorithm in which only a lognormal distribution is used, the diurnal pattern is not properly reproduced. The original GTG ensemble often tends to forecast the growth of EDR after sunrise with a time delay. We attribute this effect to the indices used, which are calibrated employing a lognormal distribution with $C_1 = -2.60 m^{2/3} s^{-1}$, more representative of stable ABLs and therefore not resulting in increased EDR forecasts with the onset of daytime convection. Also, the original GTG tends to overpredict forecast EDR values during nighttime conditions, even though the reference lognormal distribution is similar to the one used by the new LLT algorithm ($C_1 = -2.7017 m^{2/3} s^{-1}$). This behavior may be attributed to the optimization criteria used for the original GTG, which are based on minimizing the AUC for EDR $= 0.22 m^{2/3} s^{-1}$ and do not penalize overpredictions above the specified threshold.

A suite of statistical scores is presented in Table 3 to quantify the performance of the LLT algorithms. These include MAPE for the entire evaluation period as well as the daytime and nighttime relative values, and the

TABLE 2. MAPE (%) for the indices that are part of the GTG combination for CAT and MWT forecast under CBL and SBL conditions, as well as the progressive GTG-ensemble MAPE as additional m indices are incorporated, and the remapping coefficients a and b [Eq. (1)] of each individual turbulence index. Optimization was performed using EDR observations/forecasts at the BAO tower at $z = 5, 50, 100, 150$, and 300 m during 1–30 Apr 2015. GTG results used 5-h HRRR forecasts as the input NWP model.

Index	MAPE _{<i>i</i>} (%)	m	MAPE _{<i>m</i>} (%)	a (m ^{2/3} s ^{−1})	b (m ^{2/3} s ^{−1})
CAT CBL ($n = 1742$)					
PBLED R	40.246	1	40.246	26.504	1.022
–NGM2	61.095	2	40.233	17.076	0.508
MSEDR	61.758	3	39.346	21.935	0.884
GTG ensemble	—	3	39.346	—	—
CAT SBL ($n = 1370$)					
F3D	59.118	1	59.118	1.670	0.258
NCSU1	61.095	2	57.956	1.995	0.203
SCHGW	61.758	3	57.687	1.227	0.886
CTSQ	61.806	4	57.595	−0.014	0.364
NGM1	62.990	5	56.501	1.251	0.643
–NGM2	64.291	6	55.151	3.984	0.496
Ellrod2	64.442	7	54.621	2.609	0.490
GTG ensemble	—	7	54.621	—	—
MWT CBL ($n = 1742$)					
MWT1	62.910	1	62.910	0.682	0.122
MWT3	63.829	2	62.015	6.055	0.238
MWT7	64.743	3	60.446	4.430	0.307
GTG ensemble	—	3	60.446	—	—
MWT SBL ($n = 1370$)					
MWT5	55.222	1	55.222	−3.647	0.577
MWT11	55.815	2	53.686	−5.943	0.603
GTG ensemble	—	2	53.686	—	—

probability of detection (POD) for intervals of $\pm 20\%$ and $\pm 50\%$ around the target value, labeled as $\text{POD}_{0.2}$ and $\text{POD}_{0.5}$, respectively. For illustration purposes, we selected the median of the daytime and nighttime EDR observations, $0.14 \text{ m}^{2/3} \text{ s}^{-1}$ and $0.04 \text{ m}^{2/3} \text{ s}^{-1}$, respectively, and considered observations within a $\pm 15\%$ of such median EDR. The ratio of the two selected typical values is 3.5, providing a good assessment of the performance in forecasting EDR values over a wide range. The MAPE for CAT is improved by almost a factor of 2 by the new LLT algorithm, with a reduction from 101% in the current operational GTG down to 55%. Such improvement is larger during nighttime conditions, but MAPE is still reduced by a factor of 0.5 during daytime conditions. For MWT the error is reduced by $\sim 20\%$. The better performance in CBL conditions also reflects on the probability of detection, with increases in both $\text{POD}_{0.2}$ and $\text{POD}_{0.5}$ of ~ 1.5 – 2.0 , with respect to nighttime conditions for CAT and MWT combinations in both GTG algorithms. In all cases, the new GTG LLT exhibits comparable performance for CAT and MWT combinations. The original GTG CAT ensemble presents the largest errors, however, likely originating from the lower C_2 and calibration of the indices using altitudes above the ABL.

To gain further insight into the accuracy of the GTG algorithms, the absolute percentage error (APE) distributions from which MAPE values were derived are

presented in Fig. 6 for the CAT and MWT combinations. For the GTG CAT ensemble (Fig. 6a), the new LLT algorithm presents the largest probabilities for small APE values, with an exponential-like behavior. This type of distribution is desirable, in which the largest number of instances are associated with the smallest errors. In contrast, the current operational GTG has a flat distribution for $\text{APE} < 75\%$, followed by a substantial decrease for larger APEs. The cumulative APE curves are also included, to compare the performance of both GTG algorithms. A perfect algorithm will have a cumulative APE curve that is always equal to 1. The new GTG LLT forecasts (red line) are characterized by a faster growth to 1 than the current operational GTG ones (blue line), with a clear gain in performance revealed by the larger cumulative APE values all over the range of APE. For the GTG MWT ensembles presented in Fig. 6b, similar features are found, with similar cumulative APE curves for $\text{APE} > 100\%$. In particular, the original GTG MWT ensemble has an APE distribution peaking at $\sim 75\%$, whereas the new LLT algorithm forecasts result in an APE distribution that decreases the probability as APE increases in an exponential manner.

The EDR estimates from the sonic anemometers were filtered using a 10-min moving average to have a fair comparison with the turbulence forecasts from GTG. This is due to the nature of the algorithm, which assumes

FIG. 4. Time evolution of EDR forecasts with 5-h lead time from the CAT GTG ensemble for the new LLT algorithm using the HRRR model and the GTG combination from [Sharman and Pearson \(2017\)](#) that is based on the RAP model (current operational GTG), in comparison with EDR observations from sonic anemometers during the XPIA campaign ([Muñoz-Esparza et al. 2018](#)), 9–30 Mar 2015.

that energy is transferred from larger scales resolved by the underlying NWP model and cannot reproduce turbulent events that are originated at scales not explicitly resolved by the model. To illustrate better the limitations of these LLT algorithms, sonic-anemometer-derived EDR observations every 30 s are presented in [Fig. 7](#) for 12 March 2015 together with the corresponding hourly GTG LLT forecasts. The time evolution of observed EDR reveals variability at intrahour scales, which cannot be captured with GTG or any other turbulence algorithm that relies on NWP models with grid spacings of several kilometers and in which turbulence effects are fully parameterized. Nevertheless, the GTG LLT forecasts are able to satisfactorily reproduce the EDR nature of decay during nighttime and its subsequent increase with the onset of convective daytime conditions (~ 1500 – 1600 UTC), thus providing valuable information about the mean EDR behavior at low levels. This result demonstrates that parameterized turbulence algorithms are not sufficient if the interest lies not only on the average EDR level but also on the small-scale events. For these applications, a

large-eddy-simulation model coupled to a mesoscale NWP model can be used. Recent work by [Muñoz-Esparza et al. \(2017\)](#) has proven that such a modeling approach is promising. The much smaller grid spacing required by LES models (~ 10 m) precludes it from usage for operational forecast purposes at this time, however.

Observations of EDR from the BAO tower were used to evaluate the performance of the low-level GTG algorithms. Although observations from other locations are desirable, the BAO tower (decommissioned in June 2016) was the only source of ABL sonic-anemometer data required to derive EDR that was publicly available at the time when this research was carried out. We explored the use of in situ aircraft EDR estimates as an additional source of data for validation. In situ EDRs are reported in terms of maximum and median values within 1-min intervals ([Sharman et al. 2014](#)). Aircraft rapidly change altitude in the ABL because of takeoff/landing maneuvers, however, making it impossible to associate a particular reported EDR to a specific height and also mixing ABL and free-troposphere observations. Therefore, these difficulties preclude its usage in validating LLT EDR forecasts.

FIG. 5. As in Fig. 4, but for the MWT GTG combination.

Nevertheless, the new GTG algorithm has been qualitatively inspected over all of the CONUS. In Fig. 8, an example is shown of CAT (Figs. 8a,b) and MWT (Figs. 8c,d) GTG ensembles for the new LLT algorithm at a height above the surface z of approximately 168.5 m (fourth vertical grid point in the HRRR grid). These two cases are representative of daytime (Figs. 8a,c) and nighttime (Figs. 8b,d) conditions, valid at 1800 UTC 15 March 2017 and 1000 UTC 27 March 2017, respectively. The CAT ensembles reveal overall larger forecast EDR values over

the entire CONUS during daytime conditions (Fig. 8a) relative to the nighttime forecast case (Fig. 8b), consistent with the role of atmospheric stability in turbulence production. Moreover, mountain regions such as the Appalachian Mountains, Rocky Mountain range, Sierra Nevada, and Cascades display locally larger EDR values during both times of the day. This is expected because of terrain-induced turbulence effects, which appear to be captured by our GTG LLT algorithm. The MWT ensembles (Figs. 8c,d) exhibit larger values than the CAT

TABLE 3. Statistical scores (%) for the original (RAP) and new (LLT HRRR) GTG algorithms at the BAO tower at $z = 5, 50, 100, 150$, and 300 m, during 9–30 Mar 2015. These include MAPE for the entire evaluation period as well as the daytime and nighttime relative values, and the POD for intervals of $\pm 20\%$ and $\pm 50\%$ around the target value, labeled as $\text{POD}_{0.2}$ and $\text{POD}_{0.5}$, respectively. Observations within $\pm 15\%$ of two target values are considered, corresponding to the median EDR observed during daytime and nighttime conditions: 0.14 and $0.04 \text{ m}^{2/3} \text{ s}^{-1}$, respectively.

	GTG LLT HRRR CAT	GTG RAP CAT	GTG LLT HRRR MWT	GTG RAP MWT
MAPE _{all} ($n = 2462$)	55.834	101.013	66.463	74.252
MAPE _{day} ($n = 1182$)	40.669	61.684	68.974	64.395
MAPE _{night} ($n = 1280$)	69.432	134.477	64.171	82.639
POD _{0.2} ($0.14 \text{ m}^{2/3} \text{ s}^{-1}$)	44.968	30.621	43.897	1.713
POD _{0.5} ($0.14 \text{ m}^{2/3} \text{ s}^{-1}$)	82.227	65.096	86.081	35.760
POD _{0.2} ($0.04 \text{ m}^{2/3} \text{ s}^{-1}$)	26.531	20.000	28.283	17.677
POD _{0.5} ($0.04 \text{ m}^{2/3} \text{ s}^{-1}$)	63.674	49.388	53.535	49.495

FIG. 6. APE distributions from which MAPE values in Table 3 were derived for the new LLT algorithm using the HRRR model and GTG combination from Sharman and Pearson (2017) using the RAP model (current operational GTG): (a) CAT and (b) MWT GTG ensemble combinations. Solid lines represent the cumulative APE curves. Errors are calculated from hourly EDR forecasts corresponding to 9–30 Mar 2015 at the BAO tower.

ensembles at mountain regions. We speculate that such an effect could be related to the calibration performed over relatively homogeneous terrain and that complex terrain drivers will result in consistently larger EDR values. Further analysis using observations over large periods in these regions is required to confirm this hypothesis. It is also interesting that the CAT ensemble often predicts larger EDR values over the ocean during nighttime conditions than during daytime conditions, associated with the warmer water with respect to the surrounding air resulting in convective marine boundary layers during nighttime (see Figs. 8a and 8b over the Pacific Ocean region).

5. Summary and conclusions

We have proposed an improved low-level turbulence forecast algorithm and have implemented it within the Graphical Turbulence Guidance system. The LLT EDR forecasting algorithm is based upon recent observational results from Muñoz-Esparza et al. (2018) and leverages the recently proposed statistical remapping technique introduced by Sharman and Pearson (2017). The new LLT algorithm uses remapping to a lognormal distribution for nighttime (stable) conditions and to a log-Weibull distribution for daytime (convective) conditions, specific to ABL turbulence. The surface total heat flux is employed to determine the instances in which the ABL is convective ($H \geq 0$) or stable ($H < 0$). Moreover, different suites of turbulence indices are used depending upon atmospheric stability, to improve forecasting of turbulence under the different forcing mechanisms in stable and convective ABLs.

A 1-yr-long GTG calibration was performed using the WRF-based HRRR operational product, which has a

horizontal grid resolution of 3 km and runs over the CONUS, on the basis of 5-h forecasts with a latency of 1 h. Determination of the optimum CAT and MWT ensemble combinations was performed by minimizing the mean absolute percentage error to identify the suite of indices that provides the best agreement not only for large-turbulence cases but for EDR across the entire range of possible values, since smaller EDRs are relevant for applications such as unmanned aerial systems, takeoff and landing maneuvers, and wind energy. Validation using EDR observations from the BAO tower for the period of 9–30 March 2015 was carried out, coinciding with the XPJA campaign. The proposed LLT algorithm reduces the MAPE with respect to the GTG implementation described in Sharman and Pearson (2017) for CAT by a factor of ~ 2 (MAPE = 55%). For the MWT ensemble, the improvement was not as large, ~ 1.1 (MAPE = 66%), but mountain-wave activity was likely not present most of the time. Moreover, the probability of detection within a $\pm 20\%$ and $\pm 50\%$ range for characteristically small and large ABL values is improved by a factor of ~ 1.3 – 2.3 , respectively. In addition, EDR forecasts with GTG using the new LLT algorithm exhibit realistic features such as larger daytime values relative to nighttime over land and enhanced EDR in regions of complex terrain.

The current operational GTG (version 3) used as reference to evaluate the performance of the proposed LLT algorithm is based on the RAP model, with a horizontal grid spacing of ~ 13 km (~ 3.25 times as coarse as the HRRR model). Although some improvement by the HRRR model is expected because its higher resolution, we attribute the majority of the gain in performance by the new LLT algorithm to the use of different reference

FIG. 7. Time evolution of hourly EDR forecasts from the CAT GTG ensemble for the new LLT algorithm using the HRRR model (red dots) and high-frequency (30 s) EDR observations from sonic anemometers during the XPIA campaign (Muñoz-Esparza et al. 2018), 12 Mar 2015 (solid lines). Each panel corresponds to a different height, as in Figs. 4 and 5.

PDFs for daytime versus nighttime ABL conditions, together with the use of different suites of indices depending upon atmospheric stability and new ABL-specific indices. Quantification of the exact contribution from the higher-resolution NWP model will require recalibration of the current operational GTG, which is beyond the scope of the current study. Preliminary efforts to make a transition in the operational GTG at middle and upper levels to use of the HRRR model have revealed that the characteristics of the resulting EDR forecasts are somewhat similar in magnitude and structure, with an increase in granularity from the RAP model, therefore confirming our hypothesis.

A caveat of this study is the limitation of the validation to ABL observations at the BAO tower. We expect to leverage the current analysis to include observations from the recently performed Wind Forecast Improvement Project 2 (WFIP 2) field campaign in the surroundings of the Columbia River Gorge in the Pacific Northwest (<https://www.esrl.noaa.gov/gsd/renewable/wfip2.html>), and the Mountain Terrain Atmospheric

Modeling and Observations (MATERHORN) campaign in Utah (Fernando et al. 2015), thus providing turbulence data in mountainous regions. Also, in the near future, in situ EDR reports from aircrafts will provide peak time within 1-min windows, thereby allowing evaluation of GTG LLT in regions near airports over the entire CONUS. The reference PDFs used here are characteristic of relatively homogeneous ABLs. It is speculated that the presence of terrain may modify the parameters of the distributions, likely requiring a specific subset of indices/distributions for LLT in complex terrain.

Low-level turbulence forecasting in complex-terrain regions is hindered by the ability of operational NWP model forecasts to reproduce these scenarios accurately. McCaffrey et al. (2017) have recently shown that EDR estimates from a one-dimensional PBL scheme in sub-kilometer HRRR-like forecasts in complex terrain differ from radar observations in that the output from the PBL scheme appears to be more strongly modulated by

FIG. 8. Example GTG LLT forecasts that are based on 5-h WRF-HRRR forecasts, showing contours of EDR ($\text{m}^{2/3} \text{s}^{-1}$) at a height above the surface z of approximately 168.5 m (fourth vertical grid point): (a) CAT forecast valid at 1800 UTC 15 Mar 2017, (b) CAT forecast valid at 1000 UTC 27 Mar 2017, (c) MWT forecast valid at 1800 UTC 15 Mar 2017, and (d) MWT forecast valid at 1000 UTC 27 Mar 2017. The times sign denotes the location of the BAO tower ($40^{\circ}03'N$, $105^{\circ}00'W$).

the diurnal cycle of the ABL. We hypothesize that an approach like GTG, in which turbulent indices that involve spatial gradients in both horizontal and vertical directions are utilized, may provide a way to compensate for model biases and the lack of resolution when used with an appropriate climatological description of EDR that accounts for terrain effects.

Acknowledgments. This research was funded by NASA-Ames through Grant NNX15AQ66G. The authors are especially thankful to William Chan for stimulating discussions and encouragement and to Julie K. Lundquist for providing access to the sonic-anemometer data from the XPIA campaign used to calculate EDR estimates. The authors are grateful to Branko Kosović and Jeremy Sauer for providing comments on an earlier version of the manuscript. All of the simulations were performed on the Yellowstone high-performance computer (ark:/85065/d7wd3xhc) at the National Center for Atmospheric Research (NCAR), provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation.

REFERENCES

- Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, <https://doi.org/10.1175/MWR-D-15-0242.1>.
- Bougeault, P., and P. Lacarrere, 1989: Parameterization of orography-induced turbulence in a mesobeta-scale model. *Mon. Wea. Rev.*, **117**, 1872–1890, [https://doi.org/10.1175/1520-0493\(1989\)117<1872:POOITI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1872:POOITI>2.0.CO;2).
- Cho, J. Y., R. E. Newell, B. E. Anderson, J. D. Barrick, and K. L. Thornhill, 2003: Characterizations of tropospheric turbulence and stability layers from aircraft observations. *J. Geophys. Res.*, **108**, 8784, <https://doi.org/10.1029/2002JD002820>.
- 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).
- Fernando, H. J. S., and Coauthors, 2015: The MATERHORN: Unraveling the intricacies of mountain weather. *Bull. Amer. Meteor. Soc.*, **96**, 1945–1967, <https://doi.org/10.1175/BAMS-D-13-00131.1>.
- Frehlich, R., and R. Sharman, 2004: Estimates of turbulence from numerical weather prediction model output with applications to turbulence diagnosis and data assimilation. *Mon. Wea. Rev.*, **132**, 2308–2324, [https://doi.org/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 —, 2010: Climatology of velocity and temperature turbulence statistics determined from rawinsonde and ACARS/AMDAR data. *J. Appl. Meteor. Climatol.*, **49**, 1149–1169, <https://doi.org/10.1175/2010JAMC2196.1>.
- 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>.
- Holtzlag, A., and Coauthors, 2013: Stable atmospheric boundary layers and diurnal cycles: Challenges for weather and climate models. *Bull. Amer. Meteor. Soc.*, **94**, 1691–1706, <https://doi.org/10.1175/BAMS-D-11-00187.1>.
- Holzäpfel, F., N. Tchipev, and A. Stephan, 2016: Wind impact on single vortices and counterrotating vortex pairs in ground proximity. *Flow Turbul. Combust.*, **97**, 829–848, <https://doi.org/10.1007/s10494-016-9729-2>.
- Kaplan, M. L., 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>.
- Kopardekar, P., J. Rios, T. Prevot, M. Johnson, J. Jung, and J. E. Robinson III, 2016: Unmanned aircraft system traffic management (UTM) concept of operations. *16th AIAA Aviation Technology, Integration, and Operations Conf.*, Washington, DC, AIAA, AIAA 2016-3292, https://utm.arc.nasa.gov/docs/Kopardekar_2016-3292_ATIO.pdf.
- Lindborg, E., 1999: Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence? *J. Fluid Mech.*, **388**, 259–288, <https://doi.org/10.1017/S0022112099004851>.
- Lundquist, J. K., and L. Bariteau, 2015: Dissipation of turbulence in the wake of a wind turbine. *Bound.-Layer Meteor.*, **154**, 229–241, <https://doi.org/10.1007/s10546-014-9978-3>.
- , and Coauthors, 2017: Assessing state-of-the-art capabilities for probing the atmospheric boundary layer: The XPIA field campaign. *Bull. Amer. Meteor. Soc.*, **98**, 289–314, <https://doi.org/10.1175/BAMS-D-15-00151.1>.
- Marzban, C., 2004: The ROC curve and the area under it as performance measures. *Wea. Forecasting*, **19**, 1106–1114, <https://doi.org/10.1175/825.1>.
- McCaffrey, K., L. Bianco, J. Wilczak, J. Olson, and J. Kenyon, 2017: Turbulence dissipation rates in the planetary boundary layer from wind profiling radars and mesoscale numerical weather prediction models during WFIP2. *Eighth Conf. on Weather, Climate, Water and the New Energy Economy/28th Conf. on Weather Analysis and Forecasting/24th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., J3.1, <https://ams.confex.com/ams/97Annual/webprogram/Paper309313.html>.
- Moeng, C.-H., and P. P. Sullivan, 1994: A comparison of shear- and buoyancy-driven planetary boundary layer flows. *J. Atmos. Sci.*, **51**, 999–1022, [https://doi.org/10.1175/1520-0469\(1994\)051<0999:ACOSAB>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<0999:ACOSAB>2.0.CO;2).
- Muñoz-Esparza, D., J. K. Lundquist, J. A. Sauer, B. Kosović, and R. R. Linn, 2017: Coupled mesoscale-LES modeling of a diurnal cycle during the CWEX-13 field campaign: From weather to boundary-layer eddies. *J. Adv. Model. Earth Syst.*, **9**, 1572–1594, <https://doi.org/10.1002/2017MS000960>.
- , R. Sharman, and J. Lundquist, 2018: Turbulence dissipation rate in the atmospheric boundary layer: Observations and WRF mesoscale modeling during the XPIA field campaign. *Mon. Wea. Rev.*, **146**, 351–371, <https://doi.org/10.1175/MWR-D-17-0186.1>.
- Nakanishi, M., 2001: Improvement of the Mellor–Yamada turbulence closure model based on large-eddy simulation data. *Bound.-Layer Meteor.*, **99**, 349–378, <https://doi.org/10.1023/A:1018915827400>.
- , and H. Niino, 2006: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, <https://doi.org/10.1007/s10546-005-9030-8>.
- Nastrom, G., 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, [https://doi.org/10.1175/1520-0469\(1985\)042<0950:ACOWS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1985)042<0950:ACOWS>2.0.CO;2).
- Pearson, J., and R. Sharman, 2017: Prediction of energy dissipation rates for aviation turbulence. Part II: Nowcasting convective and nonconvective turbulence. *J. Appl. Meteor. Climatol.*, **56**, 339–351, <https://doi.org/10.1175/JAMC-D-16-0312.1>.
- Reap, R., 1996: Probability forecasts of clear-air-turbulence for the contiguous US. National Weather Service Office of Meteorology Tech. Procedures Bull. 430, 15 pp, https://www.weather.gov/media/mdl/TPB_430.pdf.
- Sauer, J. A., D. Muñoz-Esparza, J. M. Canfield, K. R. Costigan, R. R. Linn, and Y.-J. Kim, 2016: A large-eddy simulation study of atmospheric boundary layer influence on stratified flows over terrain. *J. Atmos. Sci.*, **73**, 2615–2632, <https://doi.org/10.1175/JAS-D-15-0282.1>.
- Schumann, U., 2012: A contrail cirrus prediction model. *Geosci. Model Dev.*, **5**, 543–580, <https://doi.org/10.5194/gmd-5-543-2012>.
- Sharman, R., and T. Lane, 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>.
- , L. 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, <https://doi.org/10.1175/JAMC-D-13-0329.1>.
- Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032, <https://doi.org/10.1175/MWR2830.1>.
- , and J. B. Klemp, 2008: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.*, **227**, 3465–3485, <https://doi.org/10.1016/j.jcp.2007.01.037>.
- , S.-H. Park, J. B. Klemp, and C. Snyder, 2014: Atmospheric kinetic energy spectra from global high-resolution nonhydrostatic simulations. *J. Atmos. Sci.*, **71**, 4369–4381, <https://doi.org/10.1175/JAS-D-14-0114.1>.
- Smalikho, I., V. Banakh, Y. Pichugina, W. Brewer, R. Banta, J. Lundquist, and N. Kelley, 2013: Lidar investigation of atmospheric effect on a wind turbine wake. *J. Atmos. Oceanic Technol.*, **30**, 2554–2570, <https://doi.org/10.1175/JTECH-D-12-00108.1>.
- Smith, C. M., and E. D. Skillingstad, 2011: Effects of inversion height and surface heat flux on downslope windstorms. *Mon. Wea. Rev.*, **139**, 3750–3764, <https://doi.org/10.1175/2011MWR3619.1>.
- Smith, T. L., S. G. Benjamin, J. M. Brown, S. Weygandt, T. Smirnova, and B. Schwartz, 2008: Convection forecasts from the hourly updated, 3-km High Resolution Rapid Refresh (HRRR) model. *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 11.1, <https://ams.confex.com/ams/pdfpapers/142055.pdf>.