Aviation Turbulence Forecasting at DWD with ICON: Methodology, Case Studies, and Verification

TOBIAS GOECKE^a AND EKATERINA MACHULSKAYA^a

^a *Deutscher Wetterdienst, Offenbach, Germany*

(Manuscript received 20 November 2019, in final form 13 April 2021)

ABSTRACT: We present a detailed evaluation of the turbulence forecast product eddy dissipation parameter (EDP) used at the Deutscher Wetterdienst (DWD). It is based on the turbulence parameterization scheme TURBDIFF, which is operational within the Icosahedral Nonhydrostatic (ICON) numerical weather prediction model used operationally by DWD. For aviation purposes, the procedure provides the cubic root of the eddy dissipation rate $\epsilon^{1/3}$ as an overall turbulence index. This quantity is a widely used measure for turbulence intensity as experienced by aircraft. The scheme includes additional sources of turbulent kinetic energy with particular relevance to aviation, which are briefly introduced. These sources describe turbulence generation by the subgrid-scale action of wake eddies, mountain waves, and convection, as well as horizontal shear as found close to fronts or the jet stream. Furthermore, we introduce a postprocessing calibration to an empirical EDR distribution, and we demonstrate the potential as well as limitations of the final EDP-based turbulence forecast by considering several case studies of typical turbulence events. Finally, we reveal the forecasting capability of this product by verifying the model results against one year of aircraft in situ EDR measurements from commercial aircraft. We find that the forecasted EDP performs favorably when compared to the Ellrod index. In particular, the turbulence signal from deep convection, which is accounted for in the EDP product, is advantageous when spatial nonlocality is allowed in the verification procedure.

KEYWORDS: Deep convection; Mountain waves; Turbulence; Forecast verification/skill; Numerical weather prediction/forecasting; Parameterization

1. Introduction

Aviation turbulence is a major cause of injuries on commercial aircraft, sometimes causing even structural damage. Furthermore, operational costs of millions of dollars as well as flight delays and problems for the air traffic management occur due to aviation turbulence [see [Sharman and Lane \(2016\)](#) for a detailed introduction into the subject]. Progress in forecasting of turbulent events is thus required by the International Civil Aviation Organization (ICAO).

A commonly accepted measure for turbulence intensity is the cubic root of the eddy dissipation rate:

$$\text{EDR} = \epsilon^{1/3}. \quad (1)$$

Here, ϵ is the rate at which turbulent kinetic energy (TKE) is dissipated at scales typically in the millimeter range. Because ϵ can be detected on board flying aircraft ([Cornman 2016](#)), and also provided by numerical weather prediction (NWP) models, ICAO suggests using it, or rather EDR as in Eq. (1), as a turbulence index. Note that the Deutscher Wetterdienst (DWD) turbulence product called eddy dissipation parameter (EDP) is formally identical to EDR in Eq. (1), i.e., $\text{EDP} = \epsilon^{1/3}$. We use the term EDP in this work if we refer specifically to this forecast product computed in the way described below.

Recent progress in the understanding and forecasting of aviation turbulence has been along several lines. The general progress to higher model resolution of operational NWP models comes with new promises but also new challenges. A major milestone is the transition to convection permitting resolutions, which opens the possibility of explicitly simulating deep convection and its dynamical implications, such as gravity waves, and yields further insight into the generation of cloud-induced turbulence (CIT; see [Barber et al. 2018, 2019](#)). Another example is the derivation of a turbulence indication from parameterized convective gravity waves in coarse models, pursued in [S.-H. Kim et al. \(2019\)](#). Further, the prediction of mountain-wave turbulence (MWT) is affected by higher resolution of the terrain data, which might induce spurious signals in mountain-wave spectra as investigated in [Park et al. \(2016\)](#) and [J.-H. Kim et al. \(2019\)](#). Another line of development uses ensemble information to enhance the forecast quality as described in [Kim et al. \(2018\)](#) and [Storer et al. \(2019\)](#) mostly for clear-air turbulence (CAT), usually referred to as in-flight bumpiness at high altitude often related to wind shear close to the jet stream. For an overview over the various sources and mechanisms of turbulence generation, see ([Sharman et al. 2012b; Lane et al. 2012; Lane and Sharman 2014; Sharman and Trier 2019; Sharman et al. 2012a](#)).

Already in 2008, DWD decided to improve the forecasts of dangerous turbulence events over Germany and to develop a

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

Corresponding author: Tobias Goecke, tobias.goecke@dwd.de

This article is licensed under a [Creative Commons Attribution 4.0 license](http://creativecommons.org/licenses/by/4.0/) (<http://creativecommons.org/licenses/by/4.0/>).

forecast product derived from a subgrid turbulence parameterization. As a result, the operational turbulence scheme TURBDIFF (turbulent diffusion; Raschendorfer 2001; Raschendorfer et al. 2003), a TKE scheme after Mellor and Yamada (1982), operational in the limited-area NWP model Consortium for Small-Scale Modeling (COSMO; Doms et al. 2011; Baldauf et al. 2011) has been considerably extended during the years from 2008 to 2010. Sources of TKE have been introduced that contribute in the upper atmosphere, such as horizontal shear, convection and subgrid-scale orography (Raschendorfer 2011). In addition to this extension of the model, a related postprocessing has been developed during this phase, which improves the quality of the final EDP product. This has already been described in Raschendorfer and Barleben (2014) and was applied operationally within COSMO from 2011 to 2016. After the introduction of ICON (Icosahedral Nonhydrostatic Model, see Zängl et al. 2015) in 2015 the EDP product has been migrated to the new modeling framework. The EDP product is available from the global ICON model with a grid spacing $\Delta x \approx 13$ km as well as from the ICON-EU nest over Europe with $\Delta x \approx 6.5$ km. The EDP forecast is generated every 6 h with 48-h lead time. At the moment the product is used by meteorological watch offices in Germany, Austria and Switzerland. Furthermore, the data delivery to airlines and further directly into the cockpit of a particular airplane has been in a test phase since 2015. Since August 2020 data are delivered into the cockpit of aircraft from Lufthansa.

The turbulence parameterization that underlies the EDP product will also be used in the ICON-D2 domain over Germany with $\Delta x \approx 2$ km. This domain is currently too small to be of big interest for aviation so this aspect has not yet been investigated. This might be done in the future, since the use of convection permitting models reveals interesting perspectives for the prediction of CIT (see, e.g., Trier and Sharman 2016; Barber et al. 2018). The individual terms from the turbulence scheme in the (also convection permitting) COSMO model have been evaluated using boundary layer observations (Goger et al. 2018).

The main focus of this work is the evaluation of the EDP turbulence forecast product from the global ICON model. We investigate several typical situations where turbulence often occurs, including mountain waves, deep convection and horizontal wind shear. These cases help to expose the behavior of the particular additional TKE source terms, which are shown to be able to represent the relevant events quite well. Furthermore, there are situations where resolved and parameterized gravity waves contribute at the same time. Finally we also shed some light on areas with potential for further improvement. The objective verification yields, apart from an overall quality measure, further insight into the contributions from the different sources of turbulence. We differentiate between those and also investigate the dependence on the spatial uncertainty. The spatial predictability of convective events, for example, turns out to be quite different from the other sources. Furthermore, we also investigate the distribution of forecasted turbulence intensity. To optimize this aspect, a calibration procedure based on observed turbulence events is established.

The paper is organized as follows. In section 2 we provide an overview on the essential aspects of the turbulence scheme TURBDIFF. Section 3 focuses on the final EDP product including postprocessing steps. In section 4 we look at some case studies to illustrate how the different turbulence sources contribute in realistic conditions where turbulence events were observed by pilots. We present a verification against objective in situ on-board measurements of turbulence in section 5, and provide conclusions in section 6.

2. Theoretical basis

Within the ICON model, the subgrid-scale turbulent motions are parameterized through the TURBDIFF turbulence scheme (Raschendorfer 2001; Raschendorfer et al. 2003; Doms et al. 2011). Details concerning the particular source terms, relevant for aviation, can be found in Raschendorfer and Barleben (2014) and Raschendorfer (2011). For the convenience of the reader, we give a summary of those aspects of the scheme that are relevant from a turbulence forecasting perspective, in particular the TKE sources in section 2a.

TURBDIFF is based on the class of the “TKE schemes,” widely used in NWP models (Mellor and Yamada 1982). Those schemes utilize one transport prognostic equation for the TKE, often in the form of the characteristic velocity scale $q := \sqrt{2 \text{ TKE}}$. A key quantity is the TKE dissipation, which is parameterized with a relaxation ansatz:

$$\varepsilon = \frac{q^3}{\alpha_M \lambda}, \quad (2)$$

where $\alpha_M = 16.6$ is a constant stemming from the expression that relates the TKE and the friction velocity in the neutrally stratified near-surface layer, characterized by the logarithmic mean-velocity profile. The length scale λ is defined after Blackadar (1962). In the following, $\{u_1, u_2, w\}$ denote the wind components, and mean quantities are denoted with bar or capital letters ($U = \bar{u}$) and deviations from the mean with primes (u').

TKE sources in the free atmosphere

Equation (2) yields the basic estimate of the eddy dissipation rate ε . However, the TKE budget equation in its parameterized form [first four terms on the right-hand side of Eq. (9)], is not completely suitable for the description of some particular important sources of turbulence in the upper troposphere. The grid spacing of the global ICON model, which is currently about 13 km, is not sufficiently fine to explicitly resolve some motions, such as quasi-organized convective structures or fine-scale nonorographic and orographically induced gravity waves. Those processes are parameterized, as is common practice, within separate parameterization schemes. In reality, different subgrid-scale processes steadily exchange kinetic energy with each other. For example, large convective eddies induce small-scale turbulence by the turbulent cascade process, thunderstorms emit vertically and horizontally propagating gravity waves, and the breaking of nonorographic and orographically induced gravity waves generates turbulence. In the ICON model, however, as well as in many other currently used NWP and climate

models, different parameterization schemes do not generally communicate with each other directly. Instead each respond individually to the resolved-scale atmospheric situation, and change the mean flow only. Because of the importance of turbulence caused by subgrid-scale phenomena for aviation, steps toward its representation were undertaken in the TURBDIFF parameterization used in the COSMO and ICON model.

In what follows, we describe how three different processes of the turbulence generation that are important for aviation are accounted for in TURBDIFF. These sources are *mountain waves*, *horizontal wind shear*, and *convection*. The convective source includes deep and shallow convection but deep moist convection is most relevant for aviation turbulence.

1) MOUNTAIN WAVES

The emission and propagation of waves generated by unresolved flow over mountains are represented in the ICON model by the separate subgrid-scale orographic (SSO) gravity wave drag parameterization after [Lott and Miller \(1997\)](#), [Miller et al. \(1989\)](#), and [Baines and Palmer \(1990\)](#). A transfer of kinetic energy from coherent motions of gravity waves to chaotic small-scale turbulence is assumed, through the mechanism of wave breaking. The interaction of the flow with topographic inhomogeneities induces downward-directed (i.e., negative) momentum stress $\tau_i = \bar{\rho} \overline{u'_i w'}$ that is transported upward by the gravity waves. As long as a wave is stable, the stress is transported without loss. At a certain height, the wave can become unstable and break according to the wave saturation mechanism ([Palmer et al. 1986](#)), where the Kelvin–Helmholtz instability causes to collapse not the entire wave, but only its top. The wave amplitude is then reduced until stability is reestablished. This causes a reduced stress transport and the deposition of a negative amount of momentum $\partial U_i / \partial t|_{\text{SSO}} = -(1/\bar{\rho})(\partial \bar{\rho} \overline{u'_i w'} / \partial z)$, decelerating the mean flow. The kinetic energy lost by the resolved flow at this height is

$$\left. \frac{\partial E}{\partial t} \right|_{\text{SSO}} = \left. \frac{\partial}{\partial t} \right|_{\text{SSO}} \left(\frac{U_1^2 + U_2^2}{2} \right) = U_1 \left. \frac{\partial U_1}{\partial t} \right|_{\text{SSO}} + U_2 \left. \frac{\partial U_2}{\partial t} \right|_{\text{SSO}}. \quad (3)$$

The kinetic energy reservoir for gravity waves, however, is not considered in ICON explicitly. In TURBDIFF, it is assumed that the spectral domain of gravity waves lies between resolved flow and turbulent motions, and a cascade of energy through the gravity wave domain without loss is assumed. This assumption means that the kinetic energy can dissipate only at small scales close to the Kolmogorov microscale, somewhere in the millimeter range, where molecular viscosity dominates. The energy taken by waves from the resolved flow is transported to the turbulent domain where it finally dissipates. These orographic effects on TKE are embodied in term 5 on the right side of Eq. (9) (section 3).

2) HORIZONTAL SHEAR

Within the usual boundary layer approximation, horizontal gradients are neglected and thus do not contribute to the production of turbulence. However, in some situations, this assumption is not valid, such as in the vicinity of upper-level fronts and in jet stream exit regions. To account for the

generation of turbulence from horizontal shear without extending the turbulence parameterization toward a fully three-dimensional scheme, an approach with intermediate complexity was chosen. A transfer of kinetic energy is assumed from quasi two-dimensional, horizontal motions of roughly grid-scale size to the scales of motions described by the turbulence parameterization. The closure used to calculate the horizontal shear source of the TKE is similar to that from [Smagorinsky \(1963\)](#), but has a two-dimensional formulation ([Raschendorfer and Barleben 2014](#); [Raschendorfer 2011](#)). It differs in that only the deviatoric part of the two-dimensional rate-of-strain tensor $S_{ij} = (1/2)(\partial U_i / \partial x_j + \partial U_j / \partial x_i)$ is used to parameterize the subgrid-stress tensor $\overline{u'_i u'_j}$ as

$$\overline{u'_i u'_j} - \frac{1}{2} q_h^2 \delta_{ij} = -Cl_h q_h (S_{ij} - \delta_{ij} S_{kk}). \quad (4)$$

Here $\overline{u'_i u'_j}$ describes stress induced by horizontal motions larger than assumed in the usual Reynolds stress tensor, and $i, j \in \{1, 2\}$ denote horizontal components. Both sides of Eq. (4) are individually trace free. The horizontal turbulent diffusivity $Cl_h q_h$ is proportional to the length and velocity scales l_h and q_h . This length scale is a portion $\alpha_h = 0.2$ of the grid scale l . Additionally, l_h is modulated by stability conditions above $z = 1000$ m according to

$$l_h = \frac{1}{\text{Ri}^{2/3}} \alpha_h l, \quad (5)$$

where $\text{Ri} = N^2/S^2$ is the gradient Richardson number. The numerator in Ri , $N^2 = (g/\Theta)\partial_z \Theta$, is the static stability, and the denominator, $S^2 = (\partial_z U_1)^2 + (\partial_z U_2)^2$, is the square of the magnitude of the vertical shear of the horizontal wind. Also, g is the acceleration due to gravity and Θ is mean potential temperature. The length scale is thus reduced the more, the stronger the atmosphere is stratified. Below 1000 m this reduction factor decreases linearly to zero at the surface. It is also bounded from above at the stratification close to neutral. An equilibrium is assumed between horizontal shear input at grid-size scales and energy transfer down the spectrum to the scales where the usual TKE is defined. Equating the horizontal shear input with the spectral transfer, parameterized analogous to Eq. (2) but with q_h and l_h , leads to

$$S_H = \overline{u'_i u'_j} \frac{\partial U_i}{\partial x_j} = l_h^2 \sqrt{C^3 \alpha_M} \left(\sqrt{\text{DEF}^2 + c_1^2 \text{DIV}^2} - c_1 \text{DIV} \right)^3 \quad (6)$$

for the horizontal shear source. Here the divergence is defined as $\text{DIV} = \partial_1 U_1 + \partial_2 U_2$, and the deformation as $\text{DEF}^2 = \text{DSH}^2 + \text{DST}^2$, including the shearing deformation $\text{DSH} = \partial_2 U_1 + \partial_1 U_2$, and the stretching deformation $\text{DST} = \partial_1 U_1 - \partial_2 U_2$. The constant C is a tuning factor and $c_1 = 1/4\sqrt{\alpha_M/C}$.

It is worth comparing this result with the Ellrod index TI ([Ellrod and Knapp 1992](#)), a well-known turbulence indicator that is used within the turbulence forecasting strategy for aviation at several meteorological centers ([Sharman et al. 2006](#); [Sharman and Lane 2016](#)). A formulation of the Ellrod index reported in [Ellrod and Knapp \(1992\)](#) is

$$\text{TI2} = S(\text{DEF} - \text{DIV}). \quad (7)$$

Note that there is some similarity between TI2 and $\sqrt[3]{S_H}$ (the contribution to the EDP). Neglecting the term $c_1^2 \text{DIV}$ and taking $c_1 = 1$, the two turbulence predictors, TI2 and $\sqrt[3]{S_H}$ would have the same horizontal shear input (DEF – DIV). Further, l_h depends on the Richardson number, which to some extent replicates the dependence of TI2 on the vertical wind shear S . Note that a turbulence closure that uses a TKE-scheme in the vertical and a Smagorinsky prescription in the horizontal is also used in other models (Hanley et al. 2015).

3) CONVECTION

Turbulence events within and in the vicinity of clouds and thunderstorms are an additional widely documented phenomenon (see Sharman et al. (2012b) and references therein). Within clouds, such events can result from the direct generation of small-scale turbulence from large convective eddies. Gravity waves, generated by thunderstorms, might also induce turbulence far away from the cloud (Sharman et al. 2012b). The removal of the convective-scale instability in the ICON model is parameterized within a separate convection mass-flux-type Tiedtke–Bechtold parameterization scheme (Tiedtke 1989; Bechtold et al. 2008). There is, however, no direct turbulence generation by convective events, since the two schemes are independent. The same is true for the nonorographic gravity waves parameterization scheme of ICON [Orr and Bechtold (2009) and chapter 3.8.7 of Prill et al. (2019)] so that the proper representation of the transport of the convective signal to some distance away from a thunderstorm by gravity waves is also not possible. But the kinetic energy of the updraft and downdraft will eventually be transferred to turbulent kinetic energy at small scales. This process is not modeled in detail here. Instead the convective buoyancy source term $\beta \overline{w'\theta'_v}|_{\text{conv}}$ is extracted from the convection parameterization scheme. This term acts as a source of large-scale convective kinetic energy. In TURBDIFF its direct transfer to TKE is assumed, and the source term is then directly used in the equation for TKE. In this source term $\beta = g/\Theta_0$ is the buoyancy parameter, Θ_0 is a reference potential temperature, and θ_v the virtual potential temperature.

The convection scheme of Bechtold et al. (2008) is formulated in terms of specific water vapor content q_v and dry static energy $s = c_p T + gz$, where T is the temperature and c_p the

specific air heat capacity. In these terms the convective buoyancy flux may approximately be rewritten as

$$\begin{aligned} \frac{g}{\Theta} \overline{w'\theta'_v} &\approx \frac{g}{T} \overline{w'T'_v} \approx \frac{g}{T} [1 + (R-1)\overline{q_v}] \overline{w'T'} + g(R-1) \overline{w'q'_v} \\ &\approx \frac{g}{T c_p} [1 + (R-1)\overline{q_v}] \overline{w's'} + g(R-1) \overline{w'q'_v}, \end{aligned} \quad (8)$$

where the subgrid-scale pressure fluctuations and the third-order correlation are neglected and the assumption is made that $s' \approx c_p T'$. Further, $R = R_v/R_d$ is the ratio of the gas constants for water vapor and dry air. In the bulk mass-flux approximation the subgrid-scale motions are assumed to belong to either bulk updraft, bulk downdraft, or the environment (approximately the gridbox mean). Thus, the fluxes in Eq. (8) are then represented in the convection scheme as $\overline{\rho w's'} = M_u(s_u - \bar{s}) + M_d(s_d - \bar{s})$ and $\overline{\rho w'q'_v} = M_u(q_{v,u} - \bar{q}_v) + M_d(q_{v,d} - \bar{q}_v)$, where the subscripts u and d refer to the values in updrafts and downdrafts, respectively, and $M_{u/d}$ are respective mass fluxes, which are computed within the mass-flux scheme from their differential equations [for details, see Bechtold et al. (2008)]. This finalizes the formulation of Eq. (8) as the convective source of the TKE. This source, however, is not yet used operationally in ICON, but is utilized solely for the EDP forecasts in the postprocessing as detailed in section 3.

After having presented the various sources of turbulence that contribute to TKE and the EDP one might wonder what processes are missing from this approach. The most obvious candidates are subgrid-scale gravity waves from nonorographic sources, such as convection or instabilities associated with the jet stream. Both type of waves have no definite sources in the model but follow a climatological background as described in Orr and Bechtold (2009). Utilizing the wave drag from these subgrid waves would thus likely have little effect in the turbulence forecast with EDP.

3. Final turbulence product

Here we give a comprehensive overview of the aviation turbulence forecasting scheme. The TKE equation used operationally in ICON is

$$\frac{\partial(q^2/2)}{\partial t} = - \underbrace{\left(\overline{u'_1 w'} \frac{\partial U_1}{\partial z} + \overline{u'_2 w'} \frac{\partial U_2}{\partial z} \right)}_{\text{vertical shear}} + \underbrace{\beta \overline{w'\theta'_v}}_{\text{buoyancy}} + \underbrace{\frac{\partial}{\partial z} \left[\alpha_T \lambda q \frac{\partial(q^2/2)}{\partial z} \right]}_{\text{turbulent transport}} - \underbrace{\frac{q^3}{\alpha_M \lambda}}_{\text{dissipation}} + \underbrace{U_1 \frac{\partial U_1}{\partial t} \Big|_{\text{SSO}} + U_2 \frac{\partial U_2}{\partial t} \Big|_{\text{SSO}}}_{\text{subgrid-scale orography}} + \underbrace{S_H}_{\text{horizontal shear}}. \quad (9)$$

The terms on the right-hand side of Eq. (9) are TKE production due to vertical wind shear, production/destruction due to buoyancy, turbulent transport, dissipation, and generation by subgrid-scale orography [Eq. (3)] and by horizontal wind shear from Eq. (6). All terms, including the last two, are used in the TKE equation in the course of a model run, so that they do not serve exclusively for aviation purposes, but are utilized for the operational weather prediction in general.

Following the spirit of the procedure developed in Raschendorfer and Barleben (2014), the convective contribution from Eq. (8) is only diagnosed and added to the EDP estimate in the postprocessing step. The TKE source due to the horizontal wind shear, S_H , is also used in the postprocessing—for the second time—in a similar way as the convective source. The reason for this is that the magnitude of this term in the upper troposphere and in the stratosphere is quite strongly reduced by the stability correction of the length

scale (5), which was done partly as a tuning measure to optimize the general model scores.¹ The final EDP product, which is thus a combination of analytical derivations and tuning measures, is specified as

$$\text{EDP} = \frac{q}{\sqrt[3]{\alpha_M \lambda}} + \sqrt[3]{Q_{\text{con}}} + a_{\text{hsh}} \sqrt[3]{S_H}, \quad (10)$$

where q is the solution of the TKE Eq. (9), and $Q_{\text{con}} = \beta w' \theta'_v|_{\text{conv}}$ is given by (8). For a_{hsh} the value of 1.0 is currently adopted, based on objective turbulence verification scores (not shown). In the postprocessing step, Eq. (10), the ansatz $\lambda = 0.23\Delta z$ is used for the length scale, where Δz is the thickness of the vertical model layers after Mason (1989) and which is essentially of the Smagorinsky type. The EDP-product definition in Eq. (10) follows the concept, already introduced in Raschendorfer and Barleben (2014), to add several diagnostic terms to the “pure” EDP as defined in Eq. (1). In contrast to the original idea we do, however, not apply a particular statistical optimization procedure to determine weights for the different terms in Eq. (10). To simplify the following discussion we will write numerical instances of e.g., EDP or $\ln \varepsilon^{1/3}$ without units, although the definitions in Eq. (10) or (1) imply units of $[\text{EDP}] = \text{m}^{2/3} \text{s}^{-1}$. The dissipation rate ε is always implicitly considered to have units of $\text{m}^2 \text{s}^{-3}$ such that all quantities are uniquely defined.

A new step, compared to the former scheme, is the transformation of the EDP product [Eq. (10)] using the climatology extracted from in situ observations on board commercial aircraft during the years 2009–14. This climatology is described in (Sharman and Pearson 2017; Sharman et al. 2014). The transformation maps the distribution of raw model output onto the desired target distribution, which is of lognormal form. If the distribution of the model EDP would be of lognormal form too, a mapping could be achieved by a transformation that is linear in $\ln(\text{EDP})$. Since this is not the case, we have to find a general mapping $g(x)$ that maps the original distribution $f(x)$ to a target distribution $f(x)$, where $x = \ln(\text{EDP})$. We determine $g(x)$ via a numerical optimization procedure, which is detailed in appendix A. The effects of this transformation will be further investigated in the next two sections.

4. Turbulence cases

We now provide specific examples that show how the TKE sources introduced in section 2a contribute to the forecasted EDP. We discuss a mountain-wave case, a convective case and a clear air turbulence case separately, followed by frontal convection. In particular we compare ICON model hindcasts to pilot reports (PIREPs; T. Keller and R. D. Sharman 2017, personal communication), for dates where significant turbulence was reported. Note that the subjectively

TABLE 1. PIREP intensities and EDR/EDP thresholds for different categories.

Turbulence category	PIREP intensity P	EDR/EDP	Color
None	Not reported	<0.15	—
Light	≤ 3	0.15–0.22	Green
Moderate	4, 5	0.22–0.35	Yellow
Severe	6	0.35–0.6	Red
Very severe	7, 8	>0.6	Blue

reported turbulence events have to be considered with some caution because of uncertainties in their timing and precise location. In this section we perform a qualitative evaluation of the EDP turbulence product. An objective verification follows in section 5. We will consider the full EDP product from Eq. (10) including the transformation introduced in the last section, and also some of the TKE source terms separately to study their influence. We use short-term hindcasts here, usually a few hours, in order to remove the general forecast uncertainty from the analysis and exemplify the turbulent processes most clearly.

To compare our EDP product to PIREPs, a matching of intensity scales has to be done, as detailed in Sharman et al. (2014). PIREP intensities are encoded as $P \in \{0, \dots, 8\}$. We take essentially the thresholds suggested in Sharman and Pearson (2017): $\text{EDP} > 0.15$ for light, $\text{EDP} > 0.22$ for moderate and $\text{EDP} > 0.35$ for severe turbulence. In addition we denote $\text{EDP} > 0.6$ as very severe events, since this additional threshold gives a better representation of the reported intensities with values of $P = 7$ and 8 and thus allows for a better comparison between PIREPs and model output. Our choice of matching the reports into these categories can be found in Table 1.

In the following we compare the model EDP to the EDR extracted from the PIREPs. Because a perfect matching in time and space reduces the number of possible data points considerably we allow a matching window. We compare PIREPs that are within ± 1 h of the model times, within altitudes of 1000 ft (1 ft ≈ 0.305 m) for horizontal sections, and within horizontal distances of 100 km for vertical cross sections.

a. Mountain-wave case

On 29 January 2016 a large number of turbulence incidents were reported by commercial aircraft pilots over the U.S. Rocky Mountains including three severe cases. In Fig. 1a the model EDP hindcast for 1800 UTC (initialized at 1200 UTC the same day) is shown together with PIREPs on flight level 320 (i.e., 32 000 ft) for the western United States. The color coding is according to Table 1. Several regions of turbulence are visible at flight level 320 (Fig. 1a). These are strongly related to the orography of the Rocky Mountains. The model is able to reproduce the reported turbulence pattern. The EDP intensities are consistent with the intensities of the PIREPs. The maximum PIREP intensity is $P = 6$ (severe), the hindcast has regions of severe turbulence fairly close to the events but also some regions of very severe turbulence can be seen.

We also show zonal and meridional cross sections (Figs. 1b,c), indicated by dashed lines and the labels A and B. They cross at

¹ The general scores, not related to aviation, concern forecasted quantities like wind, temperature, humidity, 500-hPa geopotential, radiative fluxes, and precipitation. These are verified, also but not exclusively, against surface stations and radiosondes on a daily basis in a routine verification.

FIG. 1. Forecasted EDP (color shading) and isentropes (black lines) from the ICON model over the Midwest United States at 1800 UTC 29 Jan 2016, during a mountain-wave-induced turbulence event. Pilot reports are shown as colored dots. The color coding is defined in Table 1. (a) Top view at flight level 320

(36°N, 105.5°W) where the PIREPs and model show a region of major intensity. Figure 1b is oriented perpendicular to the major ridges. Several regions can be identified. In the lower troposphere the bending of the isentropes resembles a typical supercritical flow past the ridge top and a behavior characteristic for a hydraulic jump farther downstream. Above, there exists a vertically extending gravity wave causing very severe turbulence below the tropopause. The observed severe encounter is collocated with the hindcasted turbulence region. Furthermore, there is hindcasted turbulence at the tropopause located horizontally close to PIREPs, although these PIREPs are not severe. Then there is additional turbulence hindcasted higher up from propagating waves. Figure 1c reveals structures along the ridge line at a longitude of 105.5°W. The turbulent region just discussed is visible from another perspective at a latitude of 36°N. The other severe event is evident at about 40°N (Fig. 1c), and its horizontal location is shown in Fig. 1a.

Overall, the model captures the structure of the events quite well, although the turbulence in these two regions is greater than in corresponding nearby PIREPs. This could mean that pilots successfully avoided the turbulent regions. Or the model might overestimate the event. This cannot be judged based on the available data. We now investigate the role of separate EDP sources for this particular case using zonal cross sections of the horizontal shear (Fig. 2a) and subgrid orographic (Fig. 2b) sources. These terms have each been transformed by the identical mapping also used for the full EDP product for simplicity. Both have similar structures resembling the turbulence close to the ridge top, a narrow zone of weaker midtropospheric turbulence in the lee of the ridge, and a turbulence maximum somewhat below the tropopause. This agreement indicates consistency of the different parameterizations. The intensities of individual TKE source terms cannot be compared to the full product directly since the latter emerges from the time integration of different source terms, but under the assumption of stationarity, independence of the sources, and the neglect of possible feedbacks a comparison is valid. From this comparison we conclude that the horizontal shear source is dominant, indicating that a major part of the relevant wave modes are resolved.

b. Convection case

Around 0300 UTC 9 March 2016 one severe and one very severe turbulence incident were reported, among numerous weaker turbulence reports for a convective event in the southern United States. A synoptic overview is presented in Fig. 3. During this event, there is a 500-hPa synoptic trough in the Southwest United States and a downstream ridge over Florida (Fig. 3a). The trough has a corresponding low pressure

←

(~9700 m). Cross sections in other panels are indicated by thick dashed lines. (b) West-east cross section at 36°N. (c) South-north cross section at 105.5°W. The position of the cross sections is further denoted with the points A, B, C, and D. Cloud boundaries are shown as blue lines ($q_{cl} = 10^{-6} \text{ kg kg}^{-1}$).

FIG. 2. As in Fig. 1b, but for (a) horizontal shear source term from Eqs. (10) and (6) and (b) subgrid-scale orographic source [see Eq. (3)]. Cloud boundaries are indicated by blue lines where the cloud condensate is $q_{c,i} = 10^{-6} \text{ kg kg}^{-1}$.

system that is about to cross the mountainous region close to the U.S.–Mexico border. In this situation a northward advection of warm and moist air from the Gulf of Mexico occurs east of the low pressure system. This is evidenced by the significant CAPE values in red (Figs. 3a,b). The region of high CAPE is surrounded by precipitation (Fig. 3b), which has large contributions from the convection scheme (not shown). All turbulence reports are close to the region of conditional instability and the surrounding rain showers. In particular, the two most severe events with $\text{EDP} > 0.35$ are located where warm and moist air from the Gulf of Mexico meets colder and drier air advected from the mountainous region to the west. This is where one might expect convection to be strongest, since large CAPE of up to 3600 J kg^{-1} occurs in this location.

In Fig. 3 the EDP field has been shown in a smoothed manner to prevent the plot from being too noisy. In what follows we will show the EDP gridpoint values again since those are closer to the actual model output. Farther east, severe to very severe turbulence at flight level 310 (Fig. 4a) and severe turbulence at flight level 360 (Fig. 4b) are both located close to elevated EDP areas, although the model EDP varies in coverage and is noisy. This is of course expected, especially if parameterized convection is active. A meridional cross section along 100.5°W (Fig. 4c) does not contain EDP of very severe intensity, though weaker intensities are situated close to the observed very severe turbulence at 28°N latitude. Whereas this observed very severe event is located within the troposphere, severe events are observed above the tropopause (Figs. 4c,d).

Vertical cross sections illustrate the convection (Figs. 5a,b) and horizontal shear (Fig. 5c) sources of TKE. Here, the convective contribution dominates, but the horizontal shear source (Fig. 5c) influences the severe event above the tropopause. This is apparently a secondary effect, induced through the

(parameterized) convective activity at lower levels. The convection scheme either induces a drag or resolved-scale gravity waves (indicated by the bending of isentropes), at higher levels, which could both lead to wind shear. In each of our analyzed cases, wave activity occurs above parameterized convection and the resulting steepening of the waves causes horizontal shear. Caution is needed when identifying a particular convective event as being responsible. This is due to the erratic nature of convection, far field effects of gravity waves (Sharman and Trier 2019; S.-H. Kim et al. 2019), and the spatial and temporal uncertainty of PIREPs. Note that if those waves are subgrid, the EDP would not be able to capture them, since convection is not used as a source for non-resolved gravity waves within the ICON model at the time of writing. To investigate such a case in more detail one would ideally need an ensemble of model runs and more observations, such as radar. These aspects are beyond the scope of the current paper. However, the general synoptic and mesoscale situation is well captured by the model and also the boundary between the air masses including the turbulence signals from convective events gives a consistent picture (see Fig. 3). As in the mountain-wave case, the horizontal shear term influences EDP, though it is less dominant.

c. CAT case

Also at 0300 UTC 9 March 2016, there were reports of turbulence in the northwestern United States. Both the model-forecasted EDP and PIREPs at flight level 270 have an elongated structure (Fig. 6a), and occur in a jet exit and upper-level frontal zone (Fig. 6b). Both the model and PIREPs contain severe turbulence embedded within a broad area of moderate-or-greater turbulence. The upper front is easily identified by the vertically tilted isentropes (Fig. 6b), where large model EDP agrees well with the reported severe turbulence. Of the three sources presented in section 2a, horizontal shear dominates the TKE production (not shown). Although we do not quantify the role of vertical shear [term 1 on the right-hand side of Eq. (9)], its strong influence is assumed based on the strong vertical gradients of wind speed beneath the jet axis (Fig. 6b). In summary, the relevant processes in this CAT event are mostly resolved using the 13-km horizontal grid spacing in ICON.

d. Frontal convection

Finally we look at a case of frontal convection at 1500 UTC 14 January 2016 over the southeastern United States. Horizontal cross sections at flight levels 210 (Fig. 7a) and 270 (Fig. 7b) indicate some observed severe and possibly very severe turbulence without any close counterpart in the model. Two principal explanations are possible. Either the frontal structure is misplaced in the forecast or the forecasted turbulence intensity is too weak.

Vertical cross sections (Fig. 8) illustrate the model's ability to capture the general frontal structure and associated turbulence. The single severe to very severe report near 32.5°N (Fig. 8a) and 82°W (Fig. 8b) is located underneath but relatively close to forecasted moderate-to-severe regions. This is one of four observed reports of severe-or-greater turbulence, which are either underneath or horizontally adjacent to the

FIG. 3. Synoptic situation at 0300 UTC 9 Mar 2016 over the United States: (a) overview and (b) zoom in of the box in (a). The geopotential at 500 hPa is shown as contour lines, with the numbers indicating height in kilometers. Precipitation is in blue and CAPE in red. In addition the EDP field at flight level 310 and PIREPs are shown with color coding according to Table 1.

western part of an extensive forecasted turbulent zone (Fig. 8b). The simulated clouds (blue contours) close to the turbulent zone are induced by frontal dynamics on the grid scale, and the convective source is not very strong here (not shown). We conclude that the position of the frontal structure is accurately simulated, but the forecasted turbulence too weak.

To describe the TKE production from gridscale convection, the horizontal shear of vertical velocity would be required as a source term. At the current model resolution, gridscale vertical velocities are not large enough to realistically approximate this source term. In addition the global model might miss fine-scale waves emitted from convective regions that could cause turbulence at large distances from the cloud (Lane et al. 2012; Sharman and Trier 2019). Especially beneath the anvil of the cloud, at distances of 50 km from the main convective updraft, turbulence can be identified in large-eddy simulation studies (Lane and Sharman 2014). On the other hand the model might also miss waves induced by instabilities at the jet stream or corresponding surface fronts (Plougonven and Snyder 2007; Plougonven and Zhang 2016). With horizontal grid spacing of 13 km,

the model underrepresents waves and their sources with wavelengths of less than about 100 km. Jets and fronts are not currently considered as sources of subgrid-scale (SGS) gravity waves within the ICON model.

Summarizing the case studies, the model is able to represent a large range of turbulent events adequately. We found ample evidence that the additional sources of turbulent kinetic energy, described in section 2a, capture the essential physical processes within the approximations made. Furthermore, using a transformation onto observed climatology that was introduced in section 3 and detailed in appendix A, yields raw EDP intensities comparable to PIREP intensities in many cases (not shown). This will be further investigated in the following section, where we complement these case studies with a systematic verification against objective in situ measurements.

5. Verification

In this section the objective verification of the turbulence forecasts against EDR measurements is presented.

FIG. 4. EDP from the ICON model and pilot reports over the United States at 0300 UTC 9 Mar 2016. The notation is as in Fig. 1. (a),(b) Top views on flight levels 310 and 360, respectively. (c),(d) Vertical cross sections at 100.5° and 100°W, respectively. The position of the cross sections is given by points A and B. Cloud boundaries are indicated by $q_{ci} = 10^{-6} \text{ kg kg}^{-1}$.

FIG. 5. Vertical cross sections of convective and horizontal shear contributions at 100° and 100.5°W corresponding to Fig. 4. Horizontal shear at 100.5°W is not shown because the convective source explains the signal at about 28°N alone. Point B corresponds to the point B in Fig. 4.

a. Basics

The EDR in situ measurements, used for verification, are collected by passenger aircraft over the United States for the entire year during 2015 (R. D. Sharman and G. Meymaris 2015, personal communication). Peak EDR values from 1-min sampling intervals are grouped into hourly datasets with thousands of reports each assigned to the same time. In total there are 2.5 million data points.

In the analysis of the turbulence forecasts quality, the modeled EDP can be compared to the observations pointwise (closest model grid point). However, many case studies that we have performed subjectively, show that at times some large scale structures are predicted with some spatial or temporal phase error. We thus use the “neighborhood maximum” (Schwartz and Sobash 2017), where for each measurement the maximal model EDP in a given circle is sought for comparison. The radial uncertainty scale is given in units of degrees ($1^\circ \approx 111.2 \text{ km}$, assuming a spherical Earth).

To gain insight into the role played by the different TKE source terms (section 2a), as well as the post processing steps, including the calibration procedure (see section 3), we evaluate several versions of the EDP product. We first verify the prognostic part of the EDP [first term in Eq. (10)], which contains TKE sources due to mountain waves and horizontal shear. Cumulative contributions to the EDP are successively evaluated by adding to the first term in Eq. (10), the convective source (term 2), and then also the additional horizontal shear source (term 3). Finally the full product after the transformation onto the observed climatology is considered.

ICON turbulence forecast quality is evaluated by different forecast measures. First, we use the receiver operating

FIG. 6. (a) EDP from the ICON model and pilot reports at flight level 270 over the United States at 0300 UTC 9 Mar 2016. The notation is as in Fig. 1. (b) Cross section at a fixed latitude of 43°N indicated by points A and B. Contours of horizontal wind speed are shown in blue (60, 70, 80, and 85 m s^{-1}).

characteristic (ROC). ROC curves (e.g., Wilks 2006) characterize the quality of forecasts of a continuous variable and are often used to distinguish between events of different intensity. We apply an interpolation approach in cases where the curve is incomplete (appendix B). Our ROC curves are insensitive to calibration procedures, such as mapping onto a certain climatology. Therefore, we examine frequency bias and symmetric external dependency scores, determine the benefit of calibration, and analyze forecasts of rare events.

b. Receiver operating characteristics

For the ROC curves we use the probability of detection (POD, fraction of observed “yes events” that were correctly forecasted) and the false alarm rate (FAR, fraction of observed

FIG. 7. EDP from the ICON model compared to pilot reports at 1500 UTC 14 Jan 2016 over the southeastern United States. Shown are top views on flight levels (a) 210 and (b) 270 of model EDP and pilot reports. The color coding is according to Table 1. The labels A, B, C, and D denote the positions of the cross sections in Fig. 8.

“no events” that were incorrectly forecasted as “yes”; also called probability of false detection). ROC curves are presented for spatial uncertainty scales of 0° (Fig. 9a), 0.5° (Fig. 9b), and 1.0° (Fig. 9c). The curves in Fig. 9 correspond to the different combinations of components of the full EDP product [Eq. (9)], with the Ellrod index TI2 [Eq. (7)] plotted for comparison. The index TI2 was calibrated to the EDR distribution of the verification dataset from the year 2015.

All versions of the EDP product, including the uncalibrated purely prognostic version, outperform the Ellrod index. It is evident that the inclusion of the convective TKE source has the most clearly visible positive impact on the forecast quality, which can be seen by comparing the pure prognostic (red curve) with all other EDP versions (Figs. 9b,c). The advantage of the convective term is not visible in the local evaluation (Fig. 9a), as will be investigated further below. The addition of the horizontal shear term and the remapping of the EDP do not change the ROC curve significantly. Our interpretation is that these steps do not introduce new information into the product but amount to a calibration, which leaves the ROC curve invariant (Wilks 2006). This does not, however, imply, that these procedures are irrelevant, as is discussed below.

We now investigate the influence of the spatial uncertainty scale in greater detail using the ROC area, where larger area under the curve is indicative of better forecasts. The area under the ROC curves (Fig. 10) is greater for all intermediate steps in calculating the full EDP product than for the Ellrod index, and increases with increasing uncertainty radius. The pure prognostic part has its greatest advantage over the Ellrod index at

FIG. 8. Vertical cross sections corresponding to the top views from Fig. 7, where the locations of the cross sections are marked with thick dashed lines at 33°N and 80°W. The blue contours indicate cloud boundaries ($q_{c,i} = 10^{-6} \text{ kg kg}^{-1}$). The labels A, B, C, and D can also be identified in Fig. 7.

short uncertainty scales. This we attribute to the ability of the prognostic transport TKE equation to reproduce, sufficiently accurately in space, some physical processes, which remain disregarded in the Ellrod index formulation, such as the non-stationarity and vertical transport of TKE, turbulence generation due to subgrid-scale orography, and the impact of thermal stratification. At large uncertainty scales this advantage reduces and the two forecast products become more similar. As in Fig. 9, the inclusion of the convective source in the postprocessing step has the strongest influence on the ROC curve (Fig. 10), though locally there is no advantage. Only a finite spatial uncertainty reveals considerable improvement (Fig. 10, orange versus red curve). It is tempting to identify this spatial scale of about one degree (111.2 km) as a typical spatial uncertainty of the prediction of turbulence events caused by convection in the model. This spatial uncertainty is compatible with results from the case studies of section 4. At larger spatial scales the advantage of the convective source decreases and all EDP products seem to converge toward the same value as the Ellrod index (Fig. 10). As for the ROC curves of Fig. 9, neither

FIG. 9. ROC curves for the different stages of the EDP product [Eq. (10)] and the Ellrod index [Eq. (7)] verified against a year of aircraft data. (a) Local comparison of model against observation; (b) 0.5° and (c) 1.0° spatial uncertainty allowed.

FIG. 10. The area under the ROC curves, shown in Fig. 9, as a function of spatial uncertainty.

the horizontal shear term, nor the calibration to observed climatology have a big impact for this metric.

c. Frequency bias

Since the ROC curve is insensitive to changes of the climatology of the forecast, we now consider the frequency bias (FBI) as an additional verification metric. The FBI of the Ellrod index is close to unity (Fig. 11) because it was calibrated against the verification dataset itself. In contrast, the three EDP versions employing the convective TKE source reveal different biases (Fig. 11). The nontransformed product versions overestimate the number of moderate and strong events. In particular, the EDP estimate that includes effects of horizontal shear (prog + conv + horshear) has the greatest positive bias (Fig. 11). The transformed EDP should by intention have the bias close to unity. However, since the transformation described in section 3 is based on the distribution of the EDR for the years 2009–14 and not on the verification dataset (from 2015), there is some discrepancy.

d. Rare events with SEDI

A deficiency of the ROC curve is that it is not perfectly suited for the verification of rare events. In our case, the strong turbulence events are indeed rare: the climatological frequencies in the dataset are 1.5% and 0.15% for the moderate and severe turbulence events, respectively. It is known that both POD and FAR tend to zero if the observed frequency of events (so-called base rate) tends to zero, no matter how good the forecast is (Stephenson et al. 2008). For the verification of rare events special verification parameters have been developed that do not degenerate (do not tend to trivial values) if the frequency of events tends to zero (see, e.g., Ferro and Stephenson 2011). In that study, the symmetric extremal dependence index (SEDI) is proposed:

SEDI

$$= \frac{\log(\text{FAR}) - \log(\text{POD}) - \log(1 - \text{FAR}) + \log(1 + \text{POD})}{\log(\text{FAR}) + \log(\text{POD}) + \log(1 - \text{FAR}) + \log(1 + \text{POD})}. \quad (11)$$

FIG. 11. Frequency bias of the different stages of the EDP product and of the Ellrod index as a function of the EDP threshold. The notation is as in Fig. 10.

The logarithms of POD and FAR in the formulation of SEDI afford special attention to rare events.

In Fig. 12 SEDI as a function of threshold is shown for the uncertainty radius of 1° for different forecast approaches, both calibrated (solid lines) and noncalibrated “as is” (dotted lines). The Ellrod index is shown in the calibrated form only. First, the EDP estimates from the TKE equation (red lines) outperform the Ellrod index for light to severe turbulence, but the different methods are comparable for events with severe to very severe turbulence. For the strongest turbulence, the Ellrod index outperforms all EDP formulations, though behavior of SEDI for intensities of $\text{EDP} > 0.6$ could be influenced by statistical noise.

Second, the convective TKE source (brown and orange lines in Fig. 12) improves the SEDI of the forecast, which confirms the conclusions made above with respect to the ROC curve and the ROC area. The inclusion of the horizontal windshear in the postprocessing does not improve SEDI (cf. solid orange and

FIG. 12. Symmetric extremal dependence index of the different stages of the EDP product and of the Ellrod index as a function of the EDP threshold. Dashed lines represent raw EDP data and solid lines calibrated ones.

brown lines in Fig. 12), as was also the case with the ROC curve and the ROC area, and is even slightly detrimental for the forecast quality with respect to SEDI.

Third, the comparison of the dashed (not calibrated forecasts) and solid (calibrated forecasts) lines shows that some hedging of SEDI is still possible. For instance, the calibration of the EDP computed from the TKE equation alone (red lines) unequivocally improves SEDI, while the calibration of the EDP with both the convective and wind shear sources (brown lines), in contrast, degrades SEDI (Fig. 12). The latter indicates that SEDI can reward some frequency overestimation, since it favors better POD more than it penalizes false alarms. Finally, Fig. 12 indicates greater similarity among calibrated forecasts, which in turn indicates the discrepancy in SEDI among different noncalibrated approaches is due to bias differences that may be eliminated by a calibration.

In summary, the full EDP product also provides improvements in SEDI over the Ellrod index for most of the intensity range. If the analysis is restricted to values of $EDP < 0.6$, all forecasting schemes displays the same well-defined limit toward the rare very severe or very severe turbulence events.

One may ask the question how the EDP compares to known turbulence forecast products like the Graphical Turbulence Guidance (GTG, Sharman and Pearson 2017). A comparison with the full GTG product is beyond the scope of this work. Figures 7a and 7b from Sharman and Pearson (2017), repeated here as Figs. 13a and 13b, compare the GTG product with in situ EDR reports within different altitude bands. A direct comparison of our EDP results with these GTG results is difficult because the datasets comprise different meteorological cases, and the details of turbulence event definitions and verification methods are not identical. A fully quantitative comparison is thus not possible. But Figs. 13a and 13b also include the Ellrod index as a reference, which might also be interpreted as a reference across the two different studies. Figures 13a and 13b show ROC curves for mid- and high-altitude bands, ranging from 10 000 to 20 000 ft and from 20 000 to 50 000 ft, respectively. The GTG verification for high altitude yields higher scores than for midaltitudes. The ROC curve lies also higher than the (green) EDP ROC curve in Fig. 9b (spatial uncertainty: 1°). The GTG curve for the midaltitude band, on the other hand, is rather close to the EDP ROC curve in Fig. 9b. However, our EDP verification does not distinguish between heights, which complicates comparisons with the Ellrod index and GTG. The performance of GTG, and even more notably, the Ellrod index, worsens dramatically at lower altitudes (Figs. 13a,b). In Fig. 13c are shown the ratios of POD of GTG over POD of Ellrod, for both altitudes (violet, green), and the ratio of POD of EDP over Ellrod (from Fig. 9a, local verification, blue). In Fig. 13c the altitude-dependent quality of GTG becomes quantitative, a similar behavior is likely for EDP. The advantage of EDP over Ellrod (blue curve) is similar to GTG at high altitude (green curve) for false alarm rates above 0.4. Below that value EDP shows a marked increase. This behavior should not be overinterpreted, however. In this range the curves are determined by rather strong and

FIG. 13. (a),(b) Reproduced from Sharman and Pearson (2017, their Fig. 7) showing GTG compared to Ellrod at 20 000–50 000 ft and 10 000–20 000 ft, respectively. (c) The ratio of $POD(GTG)/POD(Ellrod)$ from (b) and (c) as well as $POD(EDP)/POD(Ellrod)$ from Fig. 9a as function of FAR.

rare events. And the used method overemphasizes the region where the ROC curves are steep.

6. Summary and conclusions

In this work we evaluated the EDP turbulence forecasting scheme, that is based on the estimate of dissipation of turbulent kinetic energy.

Within four case studies we saw that additional turbulence source terms in the scheme are relevant to describe turbulence from mountain waves, deep convection, jet-stream-related shear, and frontal convection. The first three cases could be described by the model quite well, both in terms of spatial collocation and intensity. We also saw an interaction between the different sources as well as between resolved and parameterized processes. The frontal situation posed the greatest challenge for the model. This could depend on this synoptic condition, or be due to turbulence sources that are missing in the formulation, such as horizontal gradients of vertical velocity and subgrid-scale jet-induced inertia-gravity waves.

Finally, we presented a verification against one year of in situ observations obtained on board commercial aircraft. For reference, we compared results from our EDP product to the Ellrod index, which is commonly used in aviation weather forecasting. In summary, the EDP outperformed the Ellrod index. We attribute the success of EDP to consideration of subgrid-scale orographic forcing and the impact of thermal stratification. Since EDP is defined from a prognostic equation it includes nonstationary behavior and effects from vertical turbulent transport of TKE. One major advantage of EDP is obtained from inclusion of the convective source in the postprocessing. Interestingly, the impact becomes visible only if some spatial uncertainty scale (up to about 100 km.) is considered, which may be attributed to the spatial uncertainty of the prediction of convective events in the model or the absence of convectively generated subgrid-scale gravity waves. On the contrary, the nonconvective components show their advantage already at short spatial distances. But allowing some spatial uncertainty turned out to be always beneficial. Calibration to the observed climatological distribution does not show up very clearly in some standard verification measures, which is actually a design feature of those measures, like the ROC area. Nevertheless, the calibrated EDP of unity frequency bias with respect to the climatology and the enhanced impact of the horizontal shear seems desirable for the forecasters. The direct feedback from the forecasting department was also an important guide for the design of the EDP product.

Future work may progress along two different lines. For one thing, the global ICON ensemble is operational since the end of 2017 such that probabilistic forecast products are within reach. These should be designed and evaluated in a similar manner as presented in this work. Furthermore, the model physics might be improved. The most obvious processes missing are all that include subgrid-scale gravity waves from sources other than orography, such as convection and instabilities at the jet stream. And further processes that generate TKE due to horizontal shear of vertical wind. In addition, orographically

induced subgrid-scale gravity waves propagate in the present model only vertically. As mentioned before, the convective source is not yet beneficial for general forecast and is thus not active in the model prognostics, only in the postprocessing. These aspects might even gain importance in the ensemble prediction since there a coarse resolution is used (about 40 km globally).

Acknowledgments. The TKE model TURBDIFF was developed and implemented by Matthias Raschendorfer (DWD) including the extension with additional source terms. He also initiated the project for the TKE-based approach to estimate EDP together with Axel Barleben (DWD). We thank Axel Barleben for advice and support and Axel Seifert, Detlev Majewski, and Dmitrii Mironov for the careful review of the manuscript. We further thank Teddie Keller and Robert Sharman (both NCAR) for making the pilot reports for the case studies available to us.

APPENDIX A

Transformation onto a Given Target Distribution

Let $x = \ln(\text{EDP})$ be a variable that is described by the distribution function $f(x)$ with $\int_{-\infty}^{\infty} f(x) dx = 1$. Sought is a mapping g that is applied to x (as a field in space and time) such that the resulting distribution is $\tilde{f}(x)$. In particular we require that probability is conserved for every x :

$$\tilde{f}(x) dx = f[g(x)] dg. \quad (\text{A1})$$

We formulate this as a minimization problem:

$$\min_g \left\| \tilde{f} - (f \circ g) \frac{dg}{dx} \right\|_{n,m}, \quad (\text{A2})$$

where the norm is defined as

$$\|\cdot\|_{n,m} = \int_a^b dx |\cdot|^n w_m(x) w_m(x) = \frac{(x-a)^m}{(b-a)^m}. \quad (\text{A3})$$

Here $n, m \in \mathbb{N}$ are varied in order to find a balance between the overall shape of the distribution and the particular important tail toward high values. The weighting function $w_m(x)$ puts more weight toward this tail, where the degree of the overemphasis is determined by m . The order n of the integral norm (L^n) can be used to determine how much large deviations should be punished over small ones. We found $n, m = 1, 8$ to be a working compromise.

Equation (A2) is solved via the conjugate gradient method `fmin_cg` from the package in Virtanen et al. (2020) in the Python programming language. The transformation $g(x)$ is parameterized as cubic spline with ~ 30 degrees of freedom. This also yields an approximation of dg/dx needed in the evaluation of Eq. (A2). The integration in Eq. (A3) is performed on ~ 1000 Gauss-Legendre points via Gaussian quadrature. The integration limits are chosen as $a \approx -6$ and $b \approx 0.2$.

First, we determine the distribution of $f(x)$ by selecting a representative subset from the model forecasts with lead times

between 0 and 23 h for the year 2015. The target distribution $\tilde{f}$, of lognormal form is given in Sharman and Pearson (2017). Then, we solve Eq. (A2) as explained above to obtain $g(x)$ as a spline representation. This is then used to transform the original model output used before, in order to obtain the statistical distribution of the transformed field. Finally, in the forecast operations we use a polynomial representation of the transformation due to technical reasons. This is optimized to match the spline in the region of EDP values that are relevant for turbulence warnings.

APPENDIX B

Extrapolation of ROC Curves via Odds Ratio

It may be the case that the ROC curve does not extend to the lower-left corner of the POD–FAR plane if the forecasted EDP product does not cover enough of the intensity range. Similarly, in the case of the particular dataset used in the present study, some curves do not begin in the upper-right corner, because the lower limit of the measured EDP values is not 0.0, but 0.05 (all the values below this threshold mean “no reliable measurements”). In these cases we extrapolate the ROC curve to the respective corners, which is particularly important for the computation of the area under the ROC curve. We use the fact that the odds ratio,

$$OR = \frac{POD(1 - FAR)}{FAR(1 - POD)}, \quad (B1)$$

is approximately constant along the ROC curve (Ferro and Stephenson 2011), except for large thresholds, where OR tends to infinity for the forecasts better than random (Stephenson et al. 2008). That means, OR can be computed for the known moderate part of the curve and then, using this OR, the unknown part can be completed as a function $POD = f(FAR)$ according to the expression following from Eq. (B1):

$$POD = \frac{OR \times FAR}{1 + (OR - 1) \times FAR}.$$

REFERENCES

- Baines, P. G., and T. Palmer, 1990: Rationale for a new physically-based parameterization of subgrid scale orographic effects. ECMWF Tech. Memo. 169, 12 pp., <http://www.ecmwf.int/sites/default/files/elibrary/1990/7875-rationale-new-physically-based-parametrization-subgrid-scale-orographic-effects.pdf>.
- Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon. Wea. Rev.*, **139**, 3887–3905, <https://doi.org/10.1175/MWR-D-10-05013.1>.
- Barber, K. A., G. L. Mullendore, and M. J. Alexander, 2018: Out-of-cloud convective turbulence: Estimation method and impacts of model resolution. *J. Appl. Meteor. Climatol.*, **57**, 121–136, <https://doi.org/10.1175/JAMC-D-17-0174.1>.
- , W. Deierling, G. Mullendore, C. Kessinger, R. Sharman, and D. Muñoz-Esparza, 2019: Properties of convectively induced turbulence over developing oceanic convection. *Mon. Wea. Rev.*, **147**, 3429–3444, <https://doi.org/10.1175/MWR-D-18-0409.1>.
- Bechtold, P., M. Köhler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M. J. Rodwell, F. Vitart, and G. Balsamo, 2008: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Quart. J. Roy. Meteor. Soc.*, **134**, 1337–1351, <https://doi.org/10.1002/qj.289>.
- Blackadar, A. K., 1962: The vertical distribution of wind and turbulent exchange in a neutral atmosphere. *J. Geophys. Res.*, **67**, 3095–3102, <https://doi.org/10.1029/JZ067i008p03095>.
- Cornman, L. B., 2016: Airborne in situ measurements of turbulence. *Aviation Turbulence*, R. Sharman and T. Lane, Eds., Springer, 97–120.
- Doms, G., and Coauthors, 2011: A description of the nonhydrostatic regional COSMO model. Part II: Physical parametrizations. COSMO, 161 pp., <https://www.klimanavigator.eu/imperia/md/content/csc/klimanavigator/cosmophysparamtr.pdf>.
- 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).
- Ferro, C. A., and D. B. Stephenson, 2011: Extremal dependence indices: Improved verification measures for deterministic forecasts of rare binary events. *Wea. Forecasting*, **26**, 699–713, <https://doi.org/10.1175/WAF-D-10-05030.1>.
- Goger, B., M. W. Rotach, A. Gohm, O. Fuhrer, I. Stiperski, and A. A. Holtslag, 2018: The impact of three-dimensional effects on the simulation of turbulence kinetic energy in a major alpine valley. *Bound.-Layer Meteor.*, **168**, 1–27, <https://doi.org/10.1007/s10546-018-0341-y>.
- Hanley, K. E., R. S. Plant, T. H. Stein, R. J. Hogan, J. C. Nicol, H. W. Lean, C. Halliwell, and P. A. Clark, 2015: Mixing-length controls on high-resolution simulations of convective storms. *Quart. J. Roy. Meteor. Soc.*, **141**, 272–284, <https://doi.org/10.1002/qj.2356>.
- Kim, J.-H., R. Sharman, M. Strahan, J. W. Scheck, C. Bartholomew, J. C. Cheung, P. Buchanan, and N. Gait, 2018: Improvements in nonconvective aviation turbulence prediction for the World Area Forecast System. *Bull. Amer. Meteor. Soc.*, **99**, 2295–2311, <https://doi.org/10.1175/BAMS-D-17-0117.1>.
- , R. D. Sharman, S. G. Benjamin, J. M. Brown, S.-H. Park, and J. B. Klemp, 2019: Improvement of mountain-wave turbulence forecasts in NOAA’s Rapid Refresh (RAP) Model with the hybrid vertical coordinate system. *Wea. Forecasting*, **34**, 773–780, <https://doi.org/10.1175/WAF-D-18-0187.1>.
- Kim, S.-H., H.-Y. Chun, R. D. Sharman, and S. B. Trier, 2019: Development of near-cloud turbulence diagnostics based on a convective gravity wave drag parameterization. *J. Appl. Meteor. Climatol.*, **58**, 1725–1750, <https://doi.org/10.1175/JAMC-D-18-0300.1>.
- Lane, T. P., and R. D. Sharman, 2014: Intensity of thunderstorm-generated turbulence revealed by large-eddy simulation. *Geophys. Res. Lett.*, **41**, 2221–2227, <https://doi.org/10.1002/2014GL059299>.
- , —, S. B. Trier, R. G. Fovell, and J. K. Williams, 2012: Recent advances in the understanding of near-cloud turbulence. *Bull. Amer. Meteor. Soc.*, **93**, 499–515, <https://doi.org/10.1175/BAMS-D-11-00062.1>.
- Lott, F., and M. J. Miller, 1997: A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Quart. J. Roy. Meteor. Soc.*, **123**, 101–127, <https://doi.org/10.1002/qj.49712353704>.
- Mason, P. J., 1989: Large-eddy simulation of the convective atmospheric boundary layer. *J. Atmos. Sci.*, **46**, 1492–1516, [https://doi.org/10.1175/1520-0469\(1989\)046<1492:LESOTC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1989)046<1492:LESOTC>2.0.CO;2).

- Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.*, **20**, 851–875, <https://doi.org/10.1029/RG020i004p00851>.
- Miller, M., T. Palmer, and R. Swinbank, 1989: Parametrization and influence of subgridscale orography in general circulation and numerical weather prediction models. *Meteor. Atmos. Phys.*, **40**, 84–109, <https://doi.org/10.1007/BF01027469>.
- Orr, A., and P. Bechtold, 2009: Improvement in the capturing of short-range warm season orographic precipitation in the ECMWF model. *Meteor. Atmos. Phys.*, **103**, 15–23, <https://doi.org/10.1007/s00703-008-0288-5>.
- Palmer, T., G. 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, <https://doi.org/10.1002/qj.49711247406>.
- Park, S.-H., J.-H. Kim, R. D. Sharman, and J. B. Klemp, 2016: Update of upper level turbulence forecast by reducing unphysical components of topography in the numerical weather prediction model. *Geophys. Res. Lett.*, **43**, 7718–7724, <https://doi.org/10.1002/2016GL069446>.
- Plougonven, R., and C. Snyder, 2007: Inertia–gravity waves spontaneously generated by jets and fronts. Part I: Different baroclinic life cycles. *J. Atmos. Sci.*, **64**, 2502–2520, <https://doi.org/10.1175/JAS3953.1>.
- , and F. Zhang, 2016: Gravity waves generated by jets and fronts and their relevance for clear-air turbulence. *Aviation Turbulence*, R. Sharman and T. Lane, Eds., Springer, 385–406.
- Prill, F., and Coauthors, 2019: Icon model tutorial. DWD, 263 pp., https://code.mpimet.mpg.de/attachments/download/19568/ICON_tutorial_2019.pdf.
- Raschendorfer, M., 2001: The new turbulence parametrization of Im. *COSMO Newsl.*, **1**, 89–97, <http://www.cosmo-model.org>.
- , 2011: Further steps towards a scale separated turbulence scheme. *13th COSMO General Meeting in Rome 2011*, Rome, Italy, COSMO, <http://www.cosmo-model.org/content/consortium/generalMeetings/general2011/wg3a.htm>.
- , and A. Barleben, 2014: Vorhersage fluggefährdender Turbulenz und ihre Registrierung. *promet*, **39**, 23–35.
- , C. Simmer, and P. Gross, 2003: Parameterization of turbulent transport in the atmosphere. *Dynamics of Multiscale Earth Systems*, Lecture Notes in Earth Science, Vol. 97, Springer, 167–185.
- Schwartz, C. S., and R. A. Sobash, 2017: Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations. *Mon. Wea. Rev.*, **145**, 3397–3418, <https://doi.org/10.1175/MWR-D-16-0400.1>.
- Sharman, R., and T. Lane, 2016: *Aviation Turbulence*. 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>.
- , and S. Trier, 2019: Influences of gravity waves on convectively induced turbulence (CIT): A review. *Pure Appl. Geophys.*, **176**, 1923–1958, <https://doi.org/10.1007/s00024-018-1849-2>.
- , 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>.
- , J. Doyle, and M. Shapiro, 2012a: An investigation of a commercial aircraft encounter with severe clear-air turbulence over western Greenland. *J. Appl. Meteor. Climatol.*, **51**, 42–53, <https://doi.org/10.1175/JAMC-D-11-044.1>.
- , S. Trier, T. Lane, and J. Doyle, 2012b: Sources and dynamics of turbulence in the upper troposphere and lower stratosphere: A review. *Geophys. Res. Lett.*, **39**, L12803, <https://doi.org/10.1029/2012GL051996>.
- , L. 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>.
- Smagorinsky, J., 1963: General circulation experiments with the primitive equations. I. The basic experiment. *Mon. Wea. Rev.*, **91**, 99–164, [https://doi.org/10.1175/1520-0493\(1963\)091<0099:GCEWTP>2.3.CO;2](https://doi.org/10.1175/1520-0493(1963)091<0099:GCEWTP>2.3.CO;2).
- Stephenson, D., B. Casati, C. Ferro, and C. Wilson, 2008: The extreme dependency score: A non-vanishing measure for forecasts of rare events. *Meteor. Appl.*, **15**, 41–50, <https://doi.org/10.1002/met.53>.
- Storer, L. N., P. G. Gill, and P. D. Williams, 2019: Multi-model ensemble predictions of aviation turbulence. *Meteor. Appl.*, **26**, 416–428, <https://doi.org/10.1002/met.1772>.
- Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800, [https://doi.org/10.1175/1520-0493\(1989\)117<1779:ACMFSF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2).
- Trier, S. B., and R. D. Sharman, 2016: Mechanisms influencing cirrus banding and aviation turbulence near a convectively enhanced upper-level jet stream. *Mon. Wea. Rev.*, **144**, 3003–3027, <https://doi.org/10.1175/MWR-D-16-0094.1>.
- Virtanen, P., and Coauthors, 2020: SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, **17**, 261–272, <https://doi.org/10.1038/s41592-019-0686-2>.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.
- Zängl, G., D. Reinert, P. Rípodas, and M. Baldauf, 2015: The ICON (icosahedral non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quart. J. Roy. Meteor. Soc.*, **141**, 563–579, <https://doi.org/10.1002/qj.2378>.