Geosci. Model Dev., 19, 8385–8406, 2026
https://doi.org/10.5194/gmd-19-8385-2026
© Author(s) 2026. This work is distributed under
the Creative Commons Attribution 4.0 License.

Geoscientific Model Development
Open Access

Model evaluation paper

Wind and turbulence evaluation of the ICON model (icon-2026.04) using Doppler lidar observations

Maike Ahlgrimm[1,2,★] **and Eileen Päschke**[1,★]

[1]Deutscher Wetterdienst, Frankfurter Straße 135, 63067 Offenbach, Germany
[2]Hans-Ertel-Zentrum für Wetterforschung, Frankfurter Straße 135, 63067 Offenbach, Germany
[★]These authors contributed equally to this work.

Correspondence: Maike Ahlgrimm (maike.ahlgrimm@dwd.de)

Received: 18 December 2025 – Discussion started: 19 January 2026
Revised: 9 July 2026 – Accepted: 14 July 2026 – Published: 9 September 2026

Abstract. Turbulence parameterization in numerical weather prediction (NWP) models remains a challenge, particularly as resolution continues to increase. Existing turbulence evaluation methods often rely on high-resolution benchmark simulations which commonly depend on idealized assumptions and boundary conditions. An alternative evaluation method is presented here, based on Doppler lidar (DL) retrievals of wind and turbulent properties in the lower atmospheric boundary layer providing a more realistic representation of atmospheric turbulence. Modern DL retrieval methods enable the derivation of spatio-temporal and physically consistent profiles for wind and turbulence variables, including the turbulent kinetic energy (TKE), eddy dissipation rate (EDR) and turbulent length scale, which are key parameters commonly used in turbulence parameterization. These can be directly compared to their model equivalents, yielding insights into model deficiencies. The evaluation method is demonstrated here by applying it to the TKE scheme "Turbdiff" used in the ICOsahedral Nonhydrostatic (ICON) model, run with 2.1 km horizontal mesh size in the regional NWP configuration at the German Weather Service. Diagnostics for an equitable comparison of observed and simulated TKE are applied, accounting for model contributions on scales equivalent to those of the DL retrieval.

Results show that the model successfully represents the broad patterns of the boundary layer temporal evolution over a five-day summer period with convective days and stable nights. Discrepancies are mainly found at night for near-surface layers below the low level jets that form under stable conditions. Here, excessive mixing leads to an overestimate of wind speed, TKE and EDR. Errors in winds are smaller than in TKE and EDR, as expected given the higher uncertainty of parameterized turbulence. The turbulent length scale formulation, stability functions and prescribed minimum diffusivity are identified as potential candidates contributing to this model bias. The TKE diagnostics applied to the model to approximate the scales sampled by the DL retrieval also allow insights into the relative contribution from subgrid-scale and grid-scale processes and indicate that the model is able to flexibly re-partition these contributions according to the dominant scales of the flow.

Lastly, the relevance of the demonstrated model performance is illustrated for two applications: the estimation of mixing layer height from EDR for dispersion modeling, and turbulence intensity derived from TKE for applications in the wind energy sector.

1 Introduction

Advances in computing power have enabled the reliable application of atmospheric numerical models at increasingly finer horizontal resolutions. Numerical weather prediction (NWP) centers are now approaching sub-kilometer horizontal scales for routine regional simulations (Boutle et al., 2016; Joe et al., 2018; Glazer et al., 2025), thus entering the "turbulent gray zone" where larger turbulent eddies are partially resolved, while smaller-scale turbulence still requires parameterization. Numerous studies on this topic have demonstrated that traditional parameterizations developed for coarser horizontal resolutions, where turbulence

Published by Copernicus Publications on behalf of the European Geosciences Union.

is entirely subgrid-scale, become less applicable at these scales (Wyngaard, 2004; Honnert et al., 2020; Doubrawa and Muñoz-Esparza, 2020). The proper treatment of turbulence in the gray zone is therefore an urgent and open research topic (Dudhia, 2022).

To guide the further development of turbulence parameterizations, appropriate methods for model evaluation are required. Having suitable reference data is a fundamental necessity for assessing model quality. High-resolution large-eddy simulations (LES) of canonical idealized boundary layer cases (e.g. Shin and Hong, 2013) or "real", less idealized cases (e.g. Doubrawa and Muñoz-Esparza, 2020) are frequently used as a bench mark. Alternative validation methods include validation against SYNOP observations in the operational setting, and evaluation with ground-based flux measurements or remote sensors (Goger and Dipankar, 2024). The latter have also been used, often in conjunction with LES, to directly validate the turbulent kinetic energy budget (Nilsson et al., 2016a, b; Rohanizadegan et al., 2025). Doppler lidar (DL) observations in particular have proven valuable in providing estimates of vertical and horizontal winds in the atmospheric boundary layer (ABL), their distribution moments and related turbulent properties of the ABL. The latter are mostly derived from vertically pointing DL systems (Harvey et al., 2013; Manninen et al., 2018; Lothon et al., 2009; Dewani et al., 2023).

The DL retrieval method developed by Smalikho and Banakh (2017) can be used to derive several different turbulence variables simultaneously from a single scan strategy. In addition to the horizontal wind this includes turbulent kinetic energy (TKE), eddy dissipation rate (EDR), a turbulent length scale (L_V) and momentum fluxes for the lowest 600 m of the ABL. These are all properties that are either predicted or parameterized in many commonly used turbulence parameterizations in both global and mesoscale models (He et al., 2019; Zonato et al., 2022). Unlike conventional DL measurement and retrieval approaches that rely on alternating scan modes, the method by Smalikho and Banakh (2017) ensures spatio-temporal consistency by eliminating the need for sequential switching between different measurement configurations. This is particularly beneficial for model evaluation, as it enables a physically consistent assessment of the relationships between wind and turbulence characteristics without additional uncertainties introduced by differing observational sampling properties. Especially for TKE and EDR the retrieval method is well validated against in-situ tower data (Wildmann et al., 2020).

A further advantage of these observations obtained from conically scanning DL over vertically pointing DL is that they measure the three-dimensional turbulence rather than just the vertical component, enabling a representation of anisotropic turbulent structures (Banakh and Smalikho, 2019). Accounting for this anisotropy is crucial for improving the validation of NWP turbulence parameterizations, particularly in the gray-zone regime where horizontal motions become increasingly important. Thus, scanning DLs enable a more comprehensive and physically meaningful evaluation of turbulence parameterizations in high-resolution NWP models.

Since DL measurements based on conical scans only resolve a limited range of spatial and temporal scales, it is essential to ensure consistency by extracting the corresponding scales from the model in the evaluation. This approach of approximating the scales of the observational reference is in contrast to the method typically applied to LES reference data, which is filtered to match model's effective resolution (e.g. Honnert et al., 2011).

In this article, we demonstrate the value of the DL retrieval by Smalikho and Banakh (2017) for turbulence evaluation by applying it to the ICOsahedral Nonhydrostatic (ICON) model (Zängl et al., 2015) used for NWP at the German Weather Service (DWD). In Sect. 2 we describe the DL measurement and retrieval method and the model setup, followed by the method to derive the full DL-equivalent TKE from the model simulations. The evaluation of turbulent parameters is shown in Sect. 3. A more in-depth validation of the parameterized turbulent length scale can be found in Sect. 4. A case study with focus on model deficits w.r.t. excessive mixing is discussed in Sect. 5. The relevance of our results for subsequent NWP applications is explored with two examples in Sect. 6, and we conclude our discussion in Sect. 7.

2 Data and methods

2.1 Doppler lidar wind and turbulence measurements

At the Meteorological Observatory Lindenberg – Richard-Aßmann-Observatory (MOL-RAO) a Halo Photonics StreamLine (now HALO Photonics by Lumibird) Doppler lidar for simultaneous wind and turbulence profile measurements in the lowest 600 m of the ABL is installed. The instrument is sited on the Falkenberg measurement field a few kilometers away from Lindenberg in east Germany, approximately equidistant between Berlin and the Polish border. The site consists of an open grassland area also hosting a 99 m mast, which is equipped with a USA-1 omnidirectional sonic anemometer (Metek GmbH) at the 50.3 and 90.3 m levels, and various additional instruments. The surrounding area is dominated by patchwork of agricultural land and forest over generally flat terrain. The DL, manufactured in 2019 with a 180 ns pulse length and a variable focus, is positioned relatively close to the mast, with an average distance of approximately 80 m. The applied DL measurement strategy and retrieval method follows the approach proposed in Smalikho and Banakh (2017). Using a continuous conical scan mode (i.e. the measurement geometry underlying VAD techniques) with an elevation angle of 35.3° and an azimuthal resolution of $\Delta\theta = 1$–$2°$, instantaneous radial velocities (V_r) are measured along the line-of-sight of each beam direction with a spatial resolution of 48 m. The elevation angle is particularly chosen so that the contributions of the variances (σ_u^2, σ_v^2, σ_w^2) of the three orthogonal wind components (u, v, w) to the radial velocity variance ($\sigma_{V_r}^2$) are equally weighted. This allows $\sigma_{V_r}^2$ measured at different azimuth angles to be directly related to the turbulent kinetic energy ($\mathrm{TKE} = 0.5(\sigma_u^2 + \sigma_v^2 + \sigma_w^2)$, Kropfli, 1986). The high azimuthal resolution enables the structure function method to be used to estimate the EDR. Smalikho and Banakh (2017) use the latter to calculate a correction term for TKE that compensates for its systematic underestimation in DL measurements, caused by the pulse averaging effect attenuating the contribution of small-scale fluctuations to $\sigma_{V_r}^2$. One scan circle takes about 72 s such that 25 full scans are completed within a 30 min averaging time. This ensures statistical robustness of the $\sigma_{V_r}^2$ values for each azimuth angle. Note that due to the conical scan patterns the retrieved wind and turbulence variables for each measurement height represent temporal-spatial averages over differently sized circular areas with diameters ranging between about 280 m at 100 m height and 1700 m at 600 m above ground. The DL measurements therefore include not only wind fluctuations that fall into the strict definition of isotropic, shear- or buoyancy induced turbulence as frequently assumed in classical turbulence parametrization schemes, but also contain the kinetic energy of non-turbulent circulations.

When performing conical scans with such a comparably high azimuthal resolution, the number of laser pulses per radial wind measurement is limited to $N_\mathrm{a} = 2000$ due to the relation $N_\mathrm{a} = \Delta\theta f_\mathrm{p}/\omega_\mathrm{S}$ (Banakh and Smalikho, 2013). Here $\omega_\mathrm{S} = 5^\circ\,\mathrm{s}^{-1}$ and $f_\mathrm{p} = 10\,\mathrm{kHz}$ denote the angular rotation rate and pulse repetition frequency of the DL system used. This affects the signal-to-noise ratio (SNR) serving as a measure of the quality of the radial velocity measurements. As a result, the time series of V_r measurements are characterized by an increased level of noise, especially under unfavorable atmospheric conditions (e.g. low aerosol load). An appropriate noise-filtering to deal with this problem has been developed by Päschke and Detring (2023) and was applied in the lidar data analysis for the present study.

Finally, Smalikho and Banakh (2017) propose a DL based integral length scale L_V to retrieve measurements for the size of the most energetic eddies in the atmospheric turbulent flow sampled by the instrument, by combining the DL based TKE and EDR using Kolmogorov's relation, i.e.

$$L_V = c_4 \frac{\mathrm{TKE}^{3/2}}{\mathrm{EDR}} \tag{1}$$

with c_4 = 0.3796. Equation (1) originates from Richardson's view of the energy cascade in turbulent flows and from the assumption of an energy balance between energy production by the large eddies at the beginning of the cascade and energy dissipation by the smallest eddies at the end of the cascade (Richardson, 1922) – a region commonly referred to as the inertial range. It is valid for well-developed turbulence close to a statistically homogeneous, stationary and isotropic state (Mora and Obligado, 2020).

2.2 Model forcing and configuration

We illustrate our evaluation method for the operational regional configuration of ICON over Germany with 2.1 km mesh size and 65 vertical levels (hereafter D2) with forcing derived from the operational 6.5 km ICON European nest. Given the triangular grid used by ICON, the horizontal mesh size is defined as the square root of the triangular grid cell area. The vertical coordinate is stretched with 18 model levels (and layer thickness ranging from 20 to 40 m) in the lowest 600 m of the ABL. The configuration is comparable to the operational setup, except a more recent model binary ("icon-2026.04") is used, which is available from the ICON Open Source git repository (https://gitlab.dkrz.de/icon/icon-model/, last access: 2 May 2026). Simulations are initialized at 00:00 UTC from the operational analysis and integrated over 36 h. The D2 configuration uses a subgrid-scale cloud scheme, a shallow convection parameterization, and a parameterization for subgrid-scale orographic drag. An overview of the ICON model can be found in Zängl et al. (2015). Given the focus of this paper, the treatment of turbulence will be described in more detail below.

Turbulence is parameterized in ICON with the "Turbdiff" scheme (Raschendorfer, 2001; Raschendorfer et al., 2003; Doms et al., 2021; Baldauf et al., 2011) together with the "Turbtran" surface transfer scheme (Baldauf et al., 2011). The scheme is based on a 2nd order closure on level 2.5 of anisotropy according to Mellor and Yamada (1982), and thus applies the standard closure assumptions by Rotta (1951) and Kolmogorov (1962) for quasi-isotropic turbulence, so as to substitute additional pressure-correlation terms or dissipation terms in the 2nd order equations, where the latter mainly have the form of a pure equilibrium between source- and sink-terms, except the equation for TKE, which is treated prognostically. Rather than using TKE itself, Turbdiff works with the characteristic velocity scale $q := \sqrt{2\mathrm{TKE}}$. The TKE equation, reproduced here in updated form from Goecke and Machulskaya (2021, their Eq. 9), takes the form:

$$\begin{aligned}\frac{\partial(q^2/2)}{\partial t} =& \underbrace{-\sum_{ij}\overline{u_i'u_j'}\frac{\partial\overline{u_i}}{\partial x_j}}_{\text{3D resolved shear}} + \underbrace{\beta\overline{w'\theta_v'}}_{\text{buoyancy}} + \underbrace{\frac{\partial}{\partial z}\left[\alpha_T l_B q \frac{\partial(q^2/2)}{\partial z}\right]}_{\text{turbulent transport}} \\ & - \underbrace{\frac{q^3}{\alpha_M l_B}}_{\text{eddy dissipation rate}} + \underbrace{\overline{u_1}\frac{\partial\overline{u_1}}{\partial t}\bigg|_{\mathrm{SSO}} + \overline{u_2}\frac{\partial\overline{u_2}}{\partial t}\bigg|_{\mathrm{SSO}}}_{\text{subgrid-scale orography}} \\ & + \underbrace{S_\mathrm{H}}_{\text{horizontal shear}}\end{aligned} \tag{2}$$

where u_1, u_2, w are the three wind components, mean quantities and their perturbations are denoted by bars and primes,

l_B is the turbulent length scale, $\beta = g/\Theta_0$ the buoyancy parameter and $\alpha_M = 16.6$ and $\alpha_T = 0.2$. Note that the first term on the right-hand-side includes the full three-dimensional shear of the resolved winds, while in Goecke and Machulskaya (2021) only the vertical shear was included.

Turbdiff attempts to account for the fact that not all unresolved circulations strictly satisfy the assumption that turbulence be isotropic and fall within the inertial range, which is implicit in conventional turbulence closures. These types of circulations – neither fully resolved, nor part of the parameterized turbulence spectrum – will be denoted as "non-turbulent circulations", or NTC hereafter. In ICON, subgrid-scale processes such as convection and subgrid-scale orographic drag are represented by their own parameterizations. The energy transfer into the inertial range due to non-linear interactions of these NTC processes is accounted for in Turbdiff by scale-transfer terms. The term labeled "subgrid-scale orography" in Eq. (2) represents the transfer of TKE into the inertial range from gravity waves, while the term labeled "horizontal shear" represents an additional source of TKE from unresolved, organized horizontal shear associated with e.g. upper level fronts or jet stream exit regions. (The scale-transfer term from convection is not currently included.) The derivation of these terms is described in Sects. 1 and 2 of Goecke and Machulskaya (2021). Thus horizontal shear is accounted for through two different pathways, by including horizontal shear of the resolved winds and through the additional subgrid source term, making Turbdiff more suitable for use at finer mesh sizes. Goger et al. (2018) demonstrate the benefit of including the additional subgrid horizontal shear source term in the turbulence scheme for a mesh size of 1.1 km over complex terrain. While their turbulence implementation differs in other details from the operational scheme described above, the conclusions regarding the importance of including additional horizontal shear applies nevertheless.

Turbdiff uses the Blackadar (1962) vertical length scale

$$l_B = \frac{\kappa(z - z_0)}{\left(1 + \frac{\kappa(z-z_0)}{l_m}\right)} \tag{3}$$

with $l_m = \kappa \min\{D_g, L_\mathrm{p}^\mathrm{max}\}$, z_0 the roughness length and D_g the horizontal mesh size. κ denotes the von-Karman constant and $L_\mathrm{p}^\mathrm{max} = 500\,\mathrm{m}$. $L_\mathrm{p} = \min\{D_g, L_\mathrm{p}^\mathrm{max}\}$ is the turbulent peak wave length which delineates the transition from NTCs to the true, isotropic inertial range across which the above described scale-transfer takes place. The value of 500 m for $L_\mathrm{p}^\mathrm{max}$ denotes the assumed upper limit to which fully isotropic turbulence can be maintained. Together with the representation of horizontal shear, this resolution-dependence of l_B constitutes another scale-adaptive feature of Turbdiff that make it more suitable for the application into the turbulent gray-zone.

The turbulent flux parameterization is based on K-theory where turbulent exchange coefficients for momentum and heat, i.e. K_m and K_h, play a key role in coupling the SGS turbulence to the resolved-scale flow. These exchange coefficients are defined as $K_\mathrm{m} = q l_B f_m$ and $K_\mathrm{h} = q l_B f_h$. The stability functions f_m and f_h are those described in Buzzi et al. (2011). A minimum value for the exchange coefficients of $K_\mathrm{m,min} = 0.75$ and $K_\mathrm{h,min} = 0.4$ is applied to maintain some degree of mixing under stable conditions.

Given the inclusion of the additional horizontal shear terms in the TKE equation above, the exchange coefficients are suitable to be used for the representation of both vertical and horizontal diffusion processes. However, in the first-order prognostic equations of ICON, only the vertical turbulent fluxes are included to date. Therefore, these efforts to include 3D shear in the formulation of the exchange coefficients are not yet being exploited to their full potential.

2.3 Calculation of the grid-scale TKE contribution

For a meaningful comparison of simulated TKE with the observations, all TKE contributions from the model representative of the temporal and spatial scales corresponding to those of the retrieval should be considered. Following the description of the Turbdiff scheme above, the parameterized (hereafter subgrid-scale, or SGS) TKE represents the turbulence in the inertial range, including any scale-transfer contributions from NTC. These NTCs (i.e. convection, SSO drag) impact the resolved model state via their own parameterizations, but the TKE associated with them is not accounted for directly. By calculating a grid-scale (GS) contribution to the TKE, we capture the impact of resolved wind fluctuations, and also attempt to capture the contribution from the NTCs via their imprint of the on the resolved flow.

Given the Doppler lidar's elevation angle of 35.3°, the diameter of the scan circle changes with height between about 280 m at 100 m height and 1700 m at 600 m above ground. The TKE retrieval has a temporal resolution of 30 min, meaning the observed TKE convolves both temporal and spatial wind variances. Further details on the TKE retrieval can be found in Sect. 2.1.

Given the horizontal mesh size of the ICON D2 configuration (approx. 2.1 km), the scan cone falls entirely within a single grid cell, and we can neglect contributions to the GS-TKE from resolved horizontal wind variability. The GS-TKE contribution from the D2 configuration is therefore purely based on the temporal wind variance at each individual grid point over a 30 min time window. This calculation is performed during the model run using winds at every 20 s time step, applying the method proposed by Welford (1962) for calculating incremental variances. In the following explanation, this 30 min temporal variance will be denoted by σ_t^2, and the 30 min average wind at each grid point by u_t, where each of these variables can stand for any of the three wind components.

For model simulations with finer mesh size where several grid points fall within the scan diameter, the contribution from horizontal wind variance can be accounted for by calculating an additional spatial variance term from model grid points falling within the spatial range of the (height-dependent) scan diameter. σ_s^2 then represents the spatial variance of the 30 min average winds u_t across all M grid points falling within the scan diameter:

$$\sigma_s^2 = \frac{1}{M-1}\sum_{i=1}^{M}(u_{t,i} - \overline{u_t})^2 \tag{4}$$

Here,

$$\overline{u_t} = \frac{1}{M}\sum_{i=1}^{M} u_{t,i} \tag{5}$$

denotes the spatial average of the 30 min average wind u_t across all grid points within the scan circle. Following rules for variance decomposition, the total temporal-spatial variance can then be derived as the sum of these two contributions:

$$\sigma_{\text{total}}^2 = \overline{\sigma_t^2} + \sigma_s^2 \tag{6}$$

and the GS-TKE, representative of the joint temporal-spatial variability is

$$\text{TKE}_{\text{GS}} = 0.5\left(\sigma_{\text{total},u}^2 + \sigma_{\text{total},v}^2 + \sigma_{\text{total},w}^2\right) \tag{7}$$

The SGS-TKE representative of the 30 min time window is also calculated during the model run as the arithmetic mean of the instantaneous SGS-TKE values predicted at every model time step by the Turbdiff scheme within that time period. For finer mesh sizes, the 30 min mean SGS-TKE would also be averaged across all grid points falling within the scan diameter. Since the SGS-TKE is parameterized through various source and sink terms, variance decomposition does not apply. The sum of these GS and SGS are then added to compare to the observations.

The relative importance of the GS and SGS TKE contributions will be explored in more detail in Sect. 4.

3 Evaluation of ICON operational limited area configuration

For the evaluation of selected turbulence variables of the ICON turbulence scheme we focus on a five-day period from 8–12 June 2023. Daily 36 h simulations initialized at 00:00 UTC were stitched together to provide continuous data over the time period, discarding the first 12 simulation hours of each simulation from 9–12 June 2023. By stitching forecasts together at 12:00 UTC, the simulations are uninterrupted throughout the stable night time period, and the growth and decay phases of the ABL. Model results from the closest grid point to the DL measurement site are evaluated. Profiles for wind speed, wind direction and turbulence (TKE, EDR) retrieved from DL for this period are shown in Fig. 1a–d. The period chosen encompasses five days with a convective boundary layer evolution during daytime, while low level jets (LLJ) were observed during nighttime. On several days, intense convective cells and precipitation developed, some of which produced cold pools. Wind speeds (Fig. 1a) are strongest in the LLJs at night, which develop due to strong radiative cooling at the surface and a decoupling of the upper stable boundary layer from the influence of surface friction. Daytime winds are generally weak, but increase slightly in the afternoon as more momentum is mixed down from the free troposphere into the deepening ABL. Surface heating due to solar irradiance leads to convective instability during the day, and a gradually deepening, well-mixed ABL marked by higher values of TKE and EDR (Fig. 1c and d). A sharp gradient in these turbulent properties marks the top of the well-mixed ABL during the morning growth phase. The evening transition period after sunset is accompanied by a rapid decrease of TKE and EDR values throughout the lowest 600 m of the ABL. At night, higher values of TKE and EDR can only be found in a shallow, shear-driven turbulent layer below the LLJs. During the day, the observed wind and turbulence profiles considered together provide indications of cold pool events on 8 and 10 June. Depending on the case, the signature of the cold pools in the wind field is dominated either by strong directional shifts as on 8 June or by pronounced increases in the wind speed as on 10 June. With regard to the turbulence variables, cold pools are recognizable by increased TKE and a rapid collapse of the mixed layer. A concurrent reduction of temperature, similar in nature to typical cold pool signatures found in weather mast observations (Kirsch et al., 2021) is also observed (not shown). For the time period considered here, we lack the dense station network deployed two years prior during the Field Experiment on Submesoscale Spatio-Temporal Variability at Lindenberg (Hohenegger et al., 2023, FESST-VaL) to unequivocally confirm the presence of cold pools from observations alone (Kirsch et al., 2024), however radar-observed precipitation cells in the vicinity (not shown) and similarities to cases from the FESSTVaL period analyzed elsewhere (Sakradzija et al., 2025) are highly indicative.

The observed ABL characteristics are qualitatively well represented by the operational ICON model (Fig. 1e–h). Deviations are noticeable in situations where cold pool events are observed or predicted. Inspection of 2 m temperature maps (not shown) confirm the occurrence of a weaker, more distant cold pool in the model on 8 June and a stronger cold pool initiated in close proximity to the Lindenberg location on the 10th. Exact matches in terms of the timing, intensity and location of cold pools should not be expected since the development of deep convective cells and associated downdrafts is not deterministic, such that timing and location of events can vary. The night of 9 June stands out with a LLJ extending closer to the ground than observed. This results in a strong overestimation of the winds in the lowest model lev-

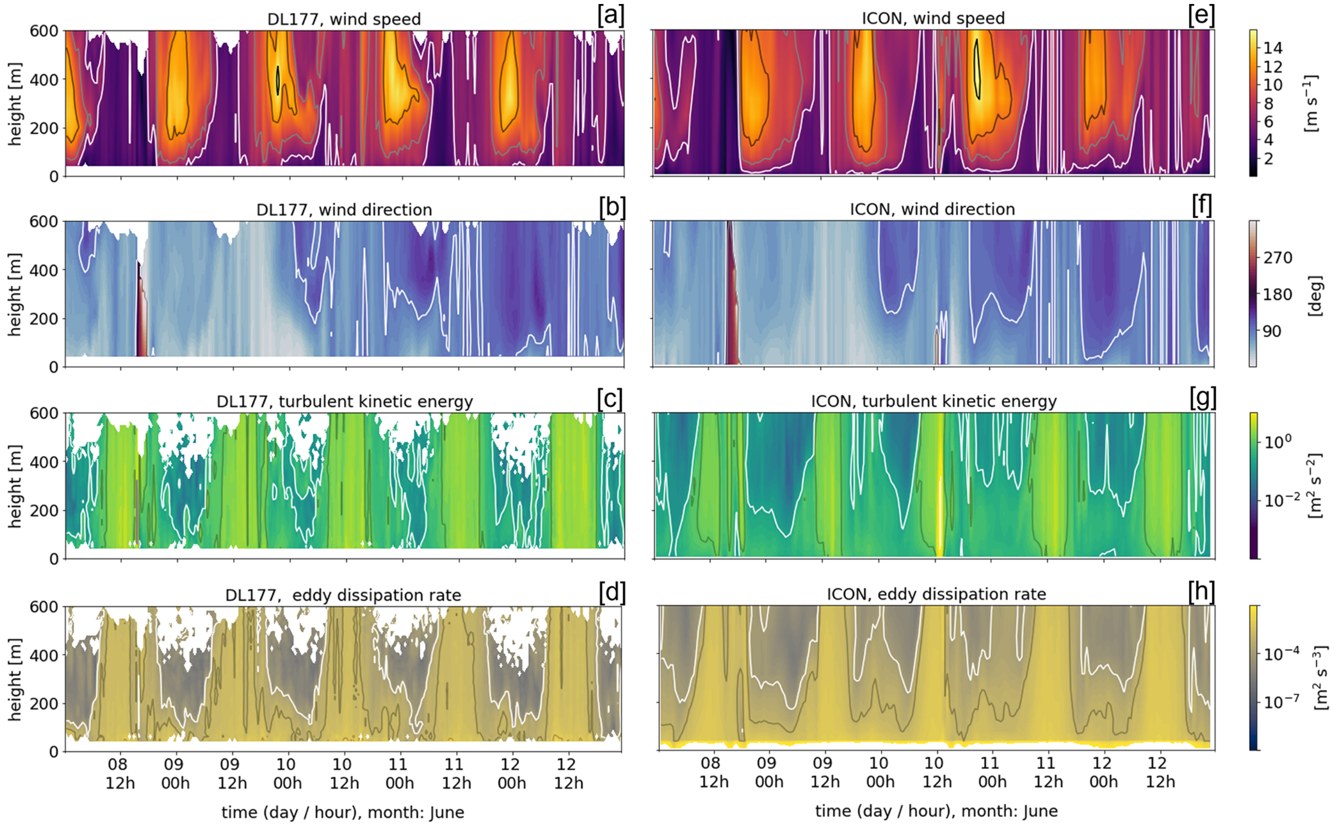

Figure 1. Comparison of time-height cross sections for **(a, e)** wind speed, **(b, f)** wind direction, **(c, g)** turbulent kinetic energy (TKE) and **(d, h)** eddy dissipation rate (EDR) from (left) DL retrieval **(a–d)** and from (right) ICON simulations **(e–h)**. Each DL-derived dataset was subject to a specific quality flag (QC), resulting in differences in data availability.

els and is also associated with a stronger turbulent mixing in ICON, which does not occur to this extent in the observations. A more detailed discussion of this case can be found in Sect. 5.

For a more systematic and quantitative evaluation of the ICON model performance, time series and time-of-day profile comparisons between ICON simulations and DL observations are shown in Figs. 2 and 3, respectively. Here, the time series plots not only provide comparisons of DL based wind and turbulence data at approximately 95 m height with corresponding ICON simulations at approximately 77 m (the closest model level height) but also include independent sonic measurements at 90 m height to further strengthen confidence in the usefulness of DL based wind and turbulence measurements for evaluating ICON during the considered time period. While the wind speed and TKE time series from the DL and sonic anemometer measurements are in very good agreement, the agreement in the EDR is weaker. The sonic based EDR estimates systematically lie between the DL observations and the model results, indicating that the DL tends to underestimate EDR relative to the sonic measurements, while the model mostly overestimates EDR, with values exceeding those observed by the sonic anemometer. Despite this bias, the present study continues to use the EDR derived from the DL, since it adequately captures the temporal evolution and relative variability of turbulence. In this sense, the EDR remains well-suited to constraining model behaviour.

The time-series comparisons between DL and ICON shown in Fig. 2 reveal quantitative differences mostly between ±30 % for wind speed, though in individual cases, such as a mis-matched cold pool event, the wind speed error can be around ±100 % (Fig. 2a). Additionally, these time series comparisons clearly show the quantitative effects of the above described erroneous LLJ simulations on 9 June, i.e. the strong overestimation of wind speeds close to the ground. For TKE (Fig. 2b), the differences between model and observation are larger and show a wider spread. In the daytime convective boundary layer, there is a tendency of the model to underestimate TKE, while at night, the model overestimates the TKE. These differences must be attributed to model deficits, as the comparison with sonic measurements confirms the high quality of the DL retrieval for wind and TKE. Note that the high quality of the DL-based TKE here provides evidence that the additional correction method in the retrieval used to compensate for the underestimation of

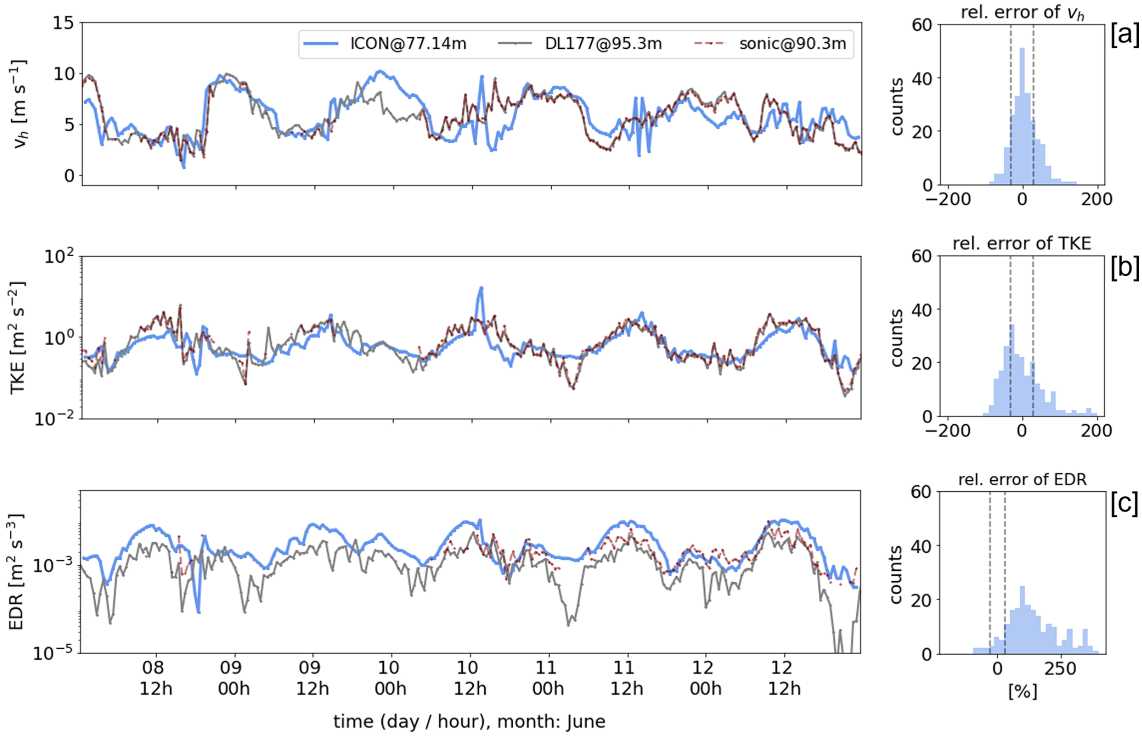

Figure 2. A five-day (8–12 June 2023) time series comparison of DL-based wind speed **(a)** and turbulence variables **(b, c)** at a height of 95.3 m with ICON simulations at the closest model height of 77.0 m. The histograms on the right show the relative errors between the two, with vertical dashed lines indicating the ±30 % range. Displayed USA-1 sonic (Metek GmbH) data at a height of 90 m operated at a distance of about 80 m from the DL system serve as an independent reference for an additional assessment of both DL and ICON data quality. Missing sonic data is due to quality control filtering (e.g. wake effects at the mast). The differing availability of TKE and EDR reflects the use of independent processing software/retrieval methods (i.e. EddyPro for TKE/Muñoz-Esparza et al., 2018 for EDR) with method-specific assumptions and quality flags. Note: The relative error for EDR is capped at 400 % in this representation to highlight the main distribution. Rare extreme outliers beyond this range were excluded from the plot to avoid scaling issues.

TKE due to pulse averaging is effective (see Sect. 2.1). In contrast, with maximum values between −100 % to +400 % the range of relative errors for EDR is considerably larger than for TKE (Fig. 2c) which, however, is inflated due to the bias in the DL estimates and so that relative errors of +400 % should be regarded as an upper bound. While correcting for this bias would reduce the discrepancy, the error is still expected to exceed that of TKE.

A quantitative intercomparison of vertical profiles up to 600 m heights between ICON and DL observations is given in Fig. 3. Median profiles and interquartile ranges (IQR) of the measured and simulated quantities at different times of day are shown, composited over the five days, to evaluate the model's ability to reproduce typical observed profiles and variability. For wind, the best profile agreement can be found during the day and at the transition from day to evening (ET) (Fig. 3a2 and a3). At night between 21:00–02:00 UTC a systematic model underestimation of the LLJ wind speed maximum in connection with a slight overestimation of the winds in the lowest parts of the profiles becomes evident (Fig. 3a4). As a consequence, the shear just below the LLJ maximum is weaker in the model when compared to the DL observations. The underestimate of the LLJ maximum disappears during the morning transition (MT) period, between 03:00–09:00 UTC (Fig. 3a1). At the same time, the LLJ features in the model profile are becoming less pronounced, pointing to an earlier LLJ decay. Similar to the wind profiles, in the TKE profiles better agreement can be observed during the day and during the ET (Fig. 3b2 and b3). By contrast, a systematic overestimation of TKE is observed at night, which changes to a systematic underestimation at the transition from night to day (Fig. 3b4 and b1). The nighttime overestimate of TKE is, at first glance, inconsistent with the weaker wind shear below the LLJ maximum in the model. A possible explanation may be differences in the stable stratification between model and observations. If the weaker LLJ in the model is concurrent with weaker stable stratification this may allow for stronger turbulent mixing despite lower wind shear. As discussed above, the apparent model overestimate of the EDR (Fig. 3c2) can be partly attributed to retrieval errors of the DL based EDR. In the evening hours, the relatively good agreement in the EDR profile especially at higher altitudes then likely indicates a model underestimate (Fig. 3c3). During the MT and daytime the model tends to overestimate the EDR (Fig. 3c1 and c2). The TKE overestimate in combination with EDR underestimate does not constitute a contradiction, as TKE and EDR characterize different aspects of turbulence. While TKE represents the total (i.e. sum of GS and SGS) turbulent kinetic energy, EDR reflects the rate at which this energy is dissipated at small scales. An underestimation of TKE combined with an overestimation of EDR therefore suggests that turbulence is too weakly generated but too rapidly dissipated in the model, implying that it does not persist long enough within the system. EDR is systematically overestimated at night, with large discrepancies even in the profile structure (Fig. 3c4). Particularly the stronger observed vertical decrease of EDR at night not reproduced by the model is striking, while the slopes of the TKE profiles are more comparable for the same time periods.

To summarize, the ICON model not only qualitatively, but also quantitatively reproduces the turbulent properties of the lower ABL well, although the results suggest that the balance between turbulence production and dissipation could be improved across different stability regimes. Additionally, the structures associated with the observed nocturnal LLJs are present, but the timing and quantitative measures, such as the LLJ maximum wind speeds and slopes of TKE and EDR profiles leave room for improvement. This is consistent with previous studies that have shown deficiencies in the representation of LLJs in the COSMO model (Consortium for Small Scale modeling, Baldauf et al., 2011), the predecessor of ICON, indicating a continuity of challenges across model generations (Finn et al., 2020). The relationship between LLJ winds, TKE and EDR will be explored in more detail in the following section and in the case studies discussed in Sect. 5.

4 Resolved and unresolved scales

The evaluation in the previous section focussed on wind, total TKE and EDR. This section analyses the contributions from the individual SGS and GS components (as discussed in Sect. 2.3) to the total TKE across the diurnal cycle. With a mesh size of 2.1 km, the ICON D2 configuration is on the threshold of the turbulent "gray zone", where large, organized turbulent structures, such as convective cells, cold pools or shear-induced instabilities associated with LLJs, are starting to be resolved to some extent at the grid scale.

A physically consistent model must exhibit scale-aware behaviour here: as soon as the grid begins to dynamically capture the most energetic components of the turbulence, the proportion of the parameterized SGS-TKE should be compensated accordingly. The aim is to evaluate whether this transition occurs successfully in diurnal variations and during transient events. This will be complemented by a comparison of the DL derived length scale L_V with the simulated counterpart, offering deeper insights into the size of the dominant eddies responsible for energy and momentum transport. Since Turbdiff relies on assumptions about the characteristic size of the most (SGS) energetic eddies, this enables a more complete understanding of the GS and SGS turbulent structures, and can be important for assessing and, if necessary, improving the Turbdiff parameterization scheme, especially under conditions with stable stratification and strong shear.

4.1 Scale partitioning of turbulent kinetic energy

The time-height plots in Fig. 4a and b show the breakdown of the total TKE as shown in Fig. 1g into its respective SGS and GS components. Both components exhibit a pronounced diurnal cycle, reflecting the daily transition of the ABL and the dynamic partitioning of energy between the two components.

To evaluate the model's ability to explicitly resolve turbulent structures in more detail, the ratio of GS-TKE to total TKE was analyzed across the diurnal cycle in Fig. 4d. When viewed in conjunction with the time-height plot for the simulated wind in Fig. 4e it is apparent that the GS-TKE contribution is strongest during the passage of the simulated cold pool on 10 June or in association with the development of the nocturnal LLJs. The briefly occurring, high GS-TKE values extending from the surface to 500 m altitude during the cold pool onset reflect the fluctuating resolved winds, and are directly followed by a collapse of the turbulent ABL, which leads to lower SGS-TKE values during the afternoon.

Finally, Fig. 4c1–c4 present a statistical comparison of median SGS-, GS-, and total TKE profiles against DL-derived measurements. This approach is employed across the four periods (MT, daytime, ET, and nighttime) previously discussed, to provide a more robust assessment of scale-adaptivity. The contrast between e.g. the daytime period, where total TKE is clearly dominated by the SGS contribution, and the evening transition period, where GS and SGS contributions are more similar in magnitude (particularly between 200–350 m altitude) shows that the model is able to flexibly partition the total TKE while maintaining overall good agreement with the observed TKE profile. The increased contribution from the GS-TKE during the evening transition composite profile is primarily driven by the development of the LLJs. The LLJ's strong wind shear creates organized structures that the model resolves directly, causing the GS-TKE component to strengthen above the increasingly stable near-surface layers. This confirms that the model successfully adapts its energy partitioning to different physical drivers.

Regarding the risk of double-counting which would manifest as a significant overestimation of total TKE compared to DL measurements, the model shows good agreement or slight underestimates during the morning and evening transitions and during daytime. An overestimate occurs only at night, and here the SGS component alone already exceeds the observed total TKE. This suggests that the bias is more

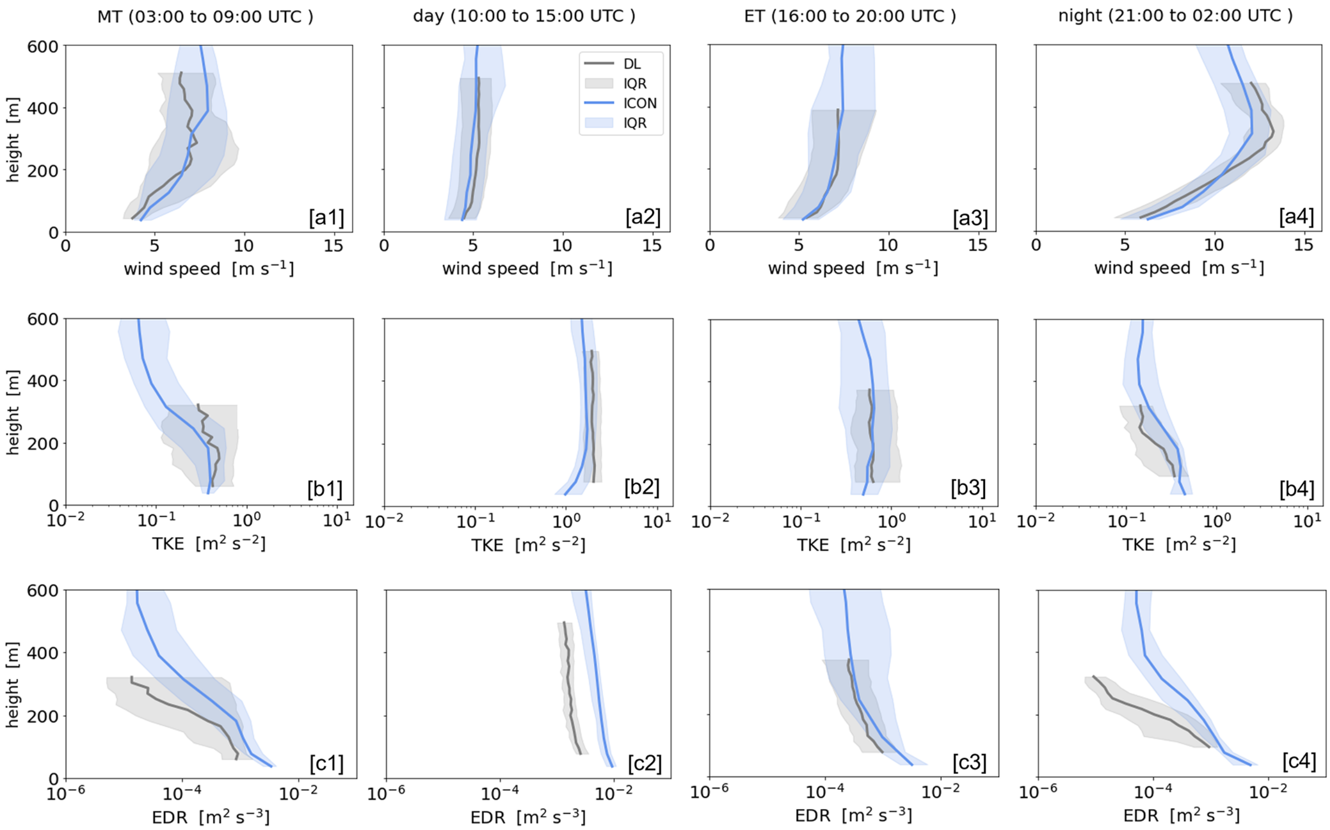

Figure 3. Comparison of a median statistics of time-of-day profiles for wind and turbulence variables for four characteristic phases (morning transition (MT), day, evening transition (ET) and night) between ICON simulations and DL observations based on the same five-day time period as shown in Fig. 1. Shaded regions denote corresponding interquartile ranges (IQR). As the measurement profiles were not always complete due to unfavourable measurement conditions in some cases, the analysis was limited to instances where observation and simulation data were available at the same time.

likely rooted in the Turbdiff parameterization scheme itself rather than being a result of double-counting.

Here, we have considered the TKE partitioning across the temporal evolution of the ABL where scale separation depends on the size of the dominant circulations relative to the (fixed) model mesh size. A corresponding analysis could be applied also to simulations across multiple grid resolutions to evaluate the scale partitioning as the mesh size changes relative to the size of the dominant circulations.

4.2 Evaluation of the turbulent length scale

As described in Sect. 2.1, a turbulent length scale L_V representing the most energetic eddies occurring within the DL scan volume can be derived from observations using Eq. (1). While the Turbdiff parameterization also utilizes a length scale l_B, this is not directly comparable to the retrieved length scale L_V since it describes the eddy size only of the SGS range of turbulence.

For a consistent comparison with DL-derived L_V, a model-derived L_V^* is reconstructed by substituting the estimation of the model's total TKE (i.e. the sum of GS and SGS contributions) and the model EDR i.e. the term "eddy dissipation rate" in Eq. (2) into Eq. (1).

Note that the EDR term in Turbdiff follows the same concept of relating TKE and EDR via a length scale as in Eq. (1), but using l_B in the model relation. While both the DL and the model relations are conceptually similar, with $c_4 = 0.3796$ and $c_\epsilon := 2^{3/2}/\alpha_M = 0.170$ differences in the employed coefficients exist. While this parameter should be a universal constant, it is in fact treated as a tuneable parameter in Turbdiff (Doms et al., 2021). Due to the scale independence of the EDR, a distinction between SGS and GS is not necessary here. Then, after some algebraic manipulation a functional relationship can be derived, reading

$$L_V^* = \frac{c_4}{c_\epsilon} l_B (1+g)^{3/2} = \frac{c_4}{c_\epsilon} l_B \left(1 + \frac{3}{2}g + \frac{3}{8}g^2 + \ldots\right) \tag{8}$$

with g defined as the ratio of GS-TKE to SGS-TKE (i.e. $g =$ GS-TKE/SGS-TKE). Note that Eq. (8) directly relates L_V^* with l_B. In the following we use L_V^* as a suitable counterpart for a comparison with L_V and therewith indirectly enable an evaluation of l_B also.

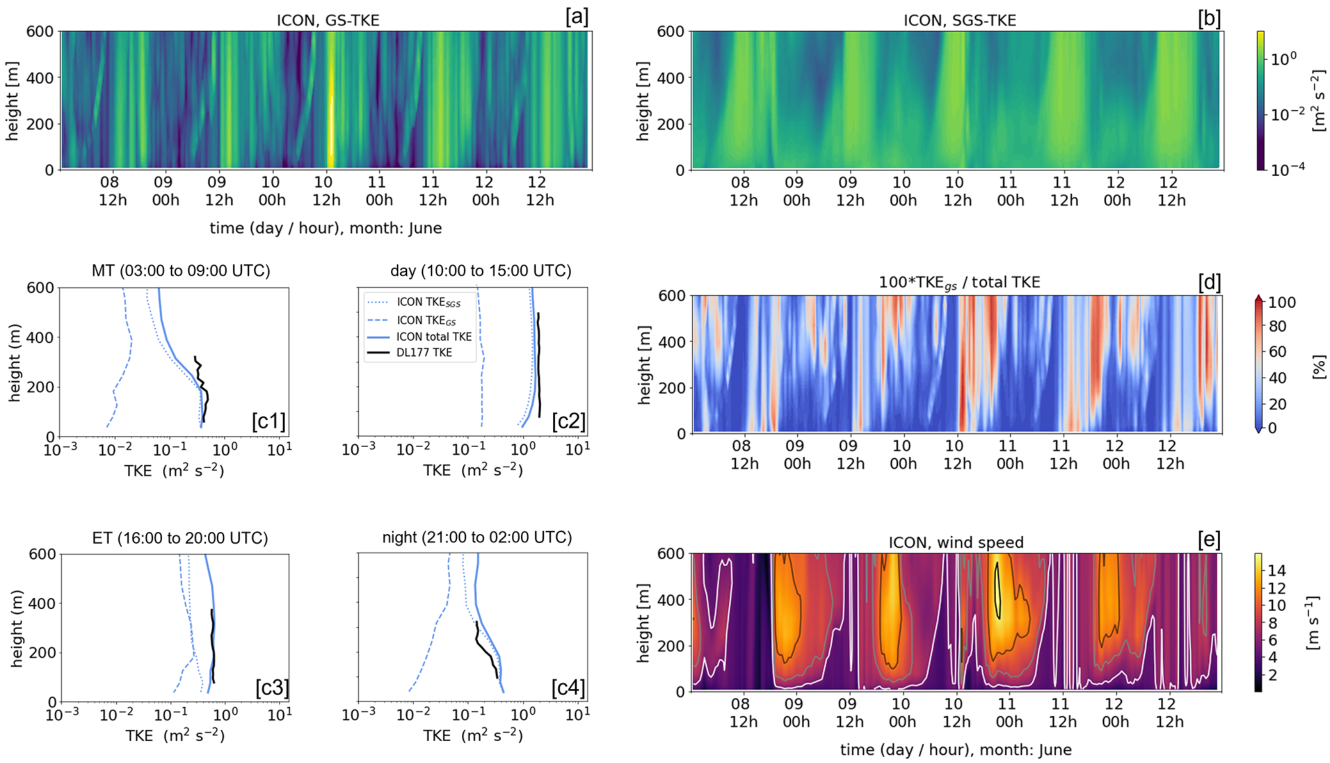

Figure 4. Time-height cross sections of **(a)** GS- and **(b)** SGS-TKE contributions, **(d)** fraction of GS-TKE relative to total TKE and **(e)** wind speed. Panels **(c1)** through **(c4)** show profiles of median (dashed) GS, (dotted) SGS and (solid) total TKE composited for the four phases of the day, as in Fig. 3.

A comparison between DL-derived L_V and L_V^* from Eq. (8) reflecting model data is shown in Fig. 5. Since L_V retrieved with Eq. (1) becomes ill-defined for very small EDR values which mainly occurs above nocturnal LLJ maxima (see Fig. 1e and h), in Fig. 5a all data are excluded where EDR < $10^{-4}\,\mathrm{m^2\,s^{-3}}$, while Fig. 5b displays all unflagged data. The time-height cross sections in Fig. 5a reveal a clear stability-dependent evolution of L_V over the course of a day. During the morning transition L_V values gradually increase from small scale (∼ 10 m) to medium-sized eddies (200–300 m). As the day progresses, L_V continues to increase and reaches maximum values of up to 1000–1500 m in the late afternoon. As the turbulence ceases towards the evening, the size of the most energetic eddies decreases again ranging between 10–200 m. The large L_V values in the late afternoon reflect typical examples of NTCs discussed in Sect. 2.2.

Considering the time-height cross sections with unflagged observational data (Fig. 5b), it can be seen in conjunction with the wind and turbulence profiles shown in Fig. 1, that in vertical regions above nocturnal LLJ maxima, where horizontal flows dominate and turbulence is weak and anisotropic, the application of Eq. (1) yields unrealistically large values for L_V. The same applies to the cold pool event on the afternoon of 8 June (see Sect. 3), whose passage involves non-stationary, rapidly evolving wind fields. This suggests that the formula is not applicable in these instances because the underlying assumptions, e.g. isotropy, quasi stationary turbulence, existence of a well-developed inertial range (see Sect. 2.1), are not fulfilled under these atmospheric conditions. This does not invalidate the relation itself, but rather delineates the limits of its applicability if used as a diagnostic for L_V from lidar measured TKE and EDR which do not stem from an inertial subrange. Note that this finding does not rule out the use of Eq. (1) to parameterize the dissipation rate of SGS-TKE in large-eddy simulations, provided that the spatial resolution is high and the cut-off wave number is indeed within the inertial range.

Figure 5c displays the model-based L_V^*, which by construction should behave similarly to L_V. It becomes obvious, however, that the patterns in the time-height cross sections for L_V^* strongly differ from the L_V observations. The comparison shows that the model does not reproduce the continuous growth of turbulent structures typically observed in a convective ABL. Significant larger-scale structures emerge only transiently, coinciding with the peak-like GS-TKE features identified in Fig. 4d. Beyond these episodes, the modeled length scales show little temporal evolution throughout the day. A closer examination of Eq. (8) and Fig. 5d illustrates that the insufficient growth of the turbulent structures directly reflects the definition of l_B. If the SGS-TKE dom-

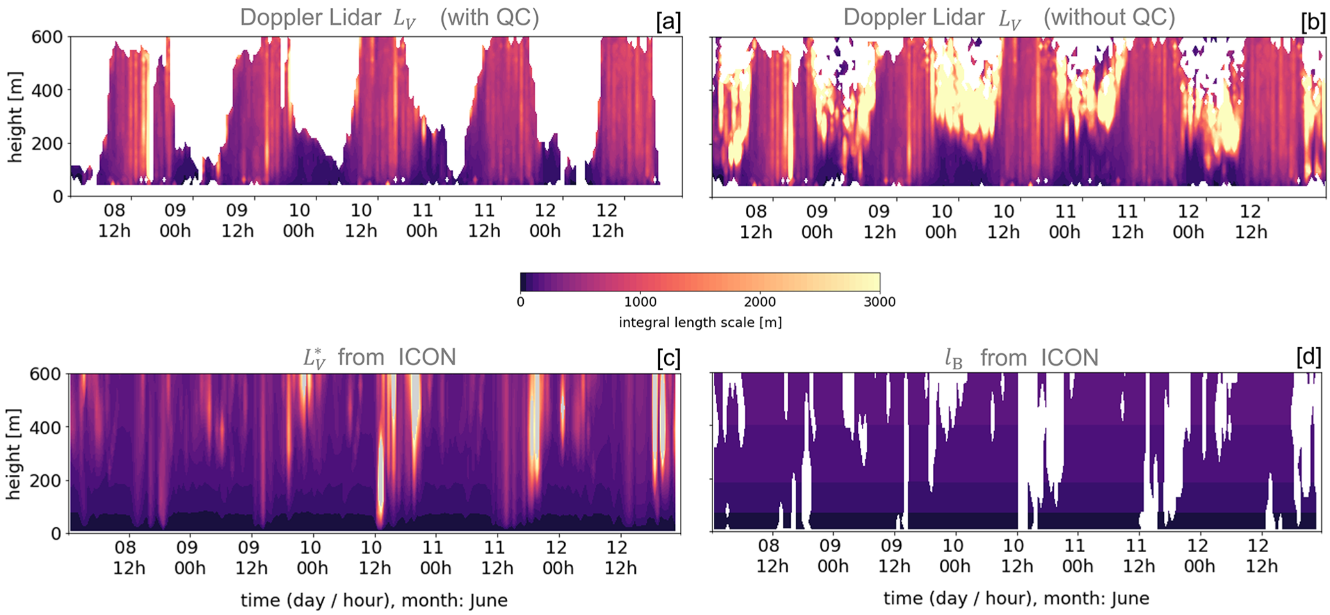

Figure 5. (a) Time-height cross sections for Doppler Lidar derived integral length scales L_V with masked area where the data fail quality control (QC), i.e. where the application of Eq. (1) is no longer justified, **(b)** same as in **(a)** but without QC, **(c)** time-height cross sections for ICON derived L_V^* calculated based on Eq. (8), **(d)** Blackadar's integral length scale l_B, which is assumed to be constant in Turbdiff, with white areas in the panel indicating where the model conditions do not satisfy $g < 1$ in Eq. (8).

inates the total TKE yielding $g < 1$, it can be shown mathematically that $L_V^* \sim c_4 c_\varepsilon^{-1} l_B$, implying that the way l_B is parameterized dictates the behaviour of L_V^* through a direct proportionality relationship. This is graphically illustrated in the similarity between Fig. 5c and d, showing L_V^* and l_B respectively. Within the context of Turbdiff, l_B is specifically constructed to exclude NTCs. One conclusion drawn from the observed differences between L_V and L_V^* might be that the TKE associated with these organized, non-isotropic circulations is not sufficiently accounted for via their parameterized imprint on the resolved flow, and our total TKE estimate is a systematic underestimate due to the missing contribution from NTCs. This might explain the missing eddies in Fig. 5c for $L_V^* > 500$ m during daytime. Yet we saw in Fig. 3 that the magnitude of the total TKE is generally in quite good agreement with the observations, particularly during the day and evening transition where larger organized circulations might be expected. Given the rather minor TKE underestimate seen for the daytime TKE, it is not likely that this argument sufficiently explains the differences in L_V and L_V^*. A second possible interpretation is that the definition of l_B based only on fixed (grid) parameters does not sufficiently reflect the effects of stability on the peak wave length $L_\mathrm{p} = l_B/\kappa$ marking the upper limit of the inertial range, and a stability-dependent formulation would better accommodate the temporal evolution of L_V^*.

5 Excessive mixing in stable boundary layers: A case study on its effects on low-level jet evolution

Building upon the preliminary finding in Sect. 3 that the strong overestimation of modeled near-surface wind speeds during the night of 9–10 June is closely related to excessive vertical mixing, this section provides a more detailed analysis of the LLJ evolution by comparing the temporal evolution of the modeled and observed profiles of wind speed and EDR in greater detail.

A comparative analysis reveals significant discrepancies between the wind profiles in Fig. 6a and b and the EDR profiles in Fig. 6c and d. DL wind speed observations show a characteristic intensification of the jet core accompanied by a distinct upward migration of the wind speed maximum. While the model does show an increase in wind speed, it produces a vertical broadening of the high-wind zone toward the surface, rather than a well-defined jet nose. This suggests that the model fails to concentrate momentum aloft, leading to the previously noted overestimation of near-surface wind speeds. The observed EDR profiles show that the air inside the jet core becomes less turbulent over time as the atmosphere stabilizes. In the model, however, the turbulence (EDR) remains high throughout the night. These persistently high EDR values imply significant vertical momentum fluxes. Instead of concentrating momentum in the jet nose, these fluxes apparently transport kinetic energy from higher altitudes towards the ground at the expense of further intensifying the LLJ maximum, which then leads to increased wind amplification below 200 m, in contrast to observations.

A slightly different perspective on the same case study is presented in Fig. 6e as a scatterplot relating vertical wind shear (S^2) and the eddy dissipation rate (EDR) from model and observations. This relationship yields an insight into the model's turbulence production efficiency. In a physically consistent stable boundary layer, high shear can exist under strongly stratified conditions without immediately triggering high dissipation (laminar regimes). In the scatterplot, we can see that the observations support such a regime of high shear (i.e. $S^2 > 10^{-4}$) with low dissipation (stable stratification), alongside a high-shear, high-dissipation (unstable stratification) regime. In the model, a much tighter relationship between these two quantities is evident, where high vertical wind shear is overwhelmingly associated with high dissipation.

This case study illustrates more clearly how the model's weak dampening of turbulence under stable conditions leads to overmixing and negatively impacts the LLJ development. To resolve the observed near-surface wind biases and the lack of atmospheric decoupling, future adjustments to the ABL scheme are required. In addition to the already discussed turbulent length scale from the previous chapter, the stability function f_B included in the calculation of the turbulent exchange coefficient is another candidate for improvement which could yield a more realistic dampening response under stable stratification. Buzzi et al. (2011) also find that "the use of a too high minimum diffusion coefficient (which is introduced in the model in order to avoid too low mixing) leads to losing important structures of the planetary boundary layer, such as the low level jet or a near-surface temperature inversion". This potential error source also applies, given that Turbdiff still uses a minimum value prescribed for K_m (see Sect. 2.2).

6 Case studies on the use of simulated turbulence variables in subsequent applications

NWP models are not solely used for forecasting weather in the narrow sense. Many of the predicted variables are essential inputs for subsequent applications such as hydrological modeling, renewable energy forecasting and air quality modeling. However, the quality of the simulations for successful follow-on applications can only be meaningfully assessed in the context of how the model results are intended to be used. In this section we want to address two particular applications that rely on wind and turbulence information. We focus on the mixing layer height (MLH) and the turbulence intensity (TI) which are diagnosed from EDR, TKE and wind model output. DWD currently post-processes daytime output from the operational ICON runs to diagnose the MLH. The diagnostic calculates the gradient Richardson number (*Ri*) and applies a critical value of 0.38 to determine the MLH. During the night (defined as the period without any insolation) a fixed MLH value is assumed, thus circumventing the issue of how to define the MLH under stable conditions. This (orography-dependent) default night-time value for the D2 configuration is 276 m for the grid point nearest the DL location. Turbulence intensity, on the other hand, is not routinely diagnosed from NWP, but may become more important in the future due to its relevance to the wind energy sector.

6.1 Mixing layer height diagnostics

The EDR time-height cross sections based on DL measurements in Fig. 1c show noticeably strong vertical gradients at certain altitudes. These gradients indicate a transition from strong turbulent mixing encountered in the ABL to the weakly turbulent free atmosphere above. As shown in Banakh et al. (2021) and O'Connor et al. (2010) this makes DL based EDR-profiles useful for the MLH determination. An independent evaluation (see Appendix A) based on the five-day period analyzed in this work confirms these findings by comparing the DL-derived MLH with radiosonde soundings (RS hereafter). In this section, ICON-based EDR profiles are used for MLH determination, in analogy with the approach used for DL data, to evaluate its potential as a supplement to traditional *Ri*-based approaches.

Figure 7a and b show time-height cross sections for the simulated virtual potential temperature (θ_v) and EDR, respectively. The additional line drawn in Fig. 7a is a result of applying the threshold approach on profiles of the bulk Richardson number (Ri_B) defined on each model level as

$$Ri_\mathrm{B} = \frac{gz(\theta_\mathrm{v} - \theta_\mathrm{s})}{\theta_\mathrm{s}(u^2 + v^2)} \tag{9}$$

where u and v are winds on the current model level, and θ_s is the surface virtual potential temperature (after Sørensen et al., 1996). A critical Richardson number of 0.2 is used, and the MLH is linearly interpolated between model level heights bracketing the critical Richardson number. This diagnostic (offline diagnostic hereafter), while conceptually similar to the operational postprocessing, is in fact different code. We adopt here the bulk Richardson number and critical value used in the MLH diagnostics applied to the radiosonde profiles (Beyrich and Leps, 2012) for consistency. In analogy with the original post-processing, during night the MLH is set to 276 m. Unlike in Fig. 1h, the time-height EDR cross section in Fig. 7b, extends up to 4000 m, showing the MLH evolution over the full day. The white contour in Fig. 7b marks the diagnosed MLH using the EDR-threshold approach described in Appendix A.

Panels (c1)–(c2) in Fig. 7 then show the respective model diagnostics together with the MLH retrieved from RS and DL measurements. It is evident that the threshold method does not always work reliably during daytime conditions, where EDR occasionally drops below the threshold at unrealistically high altitudes. During the night, the EDR-based

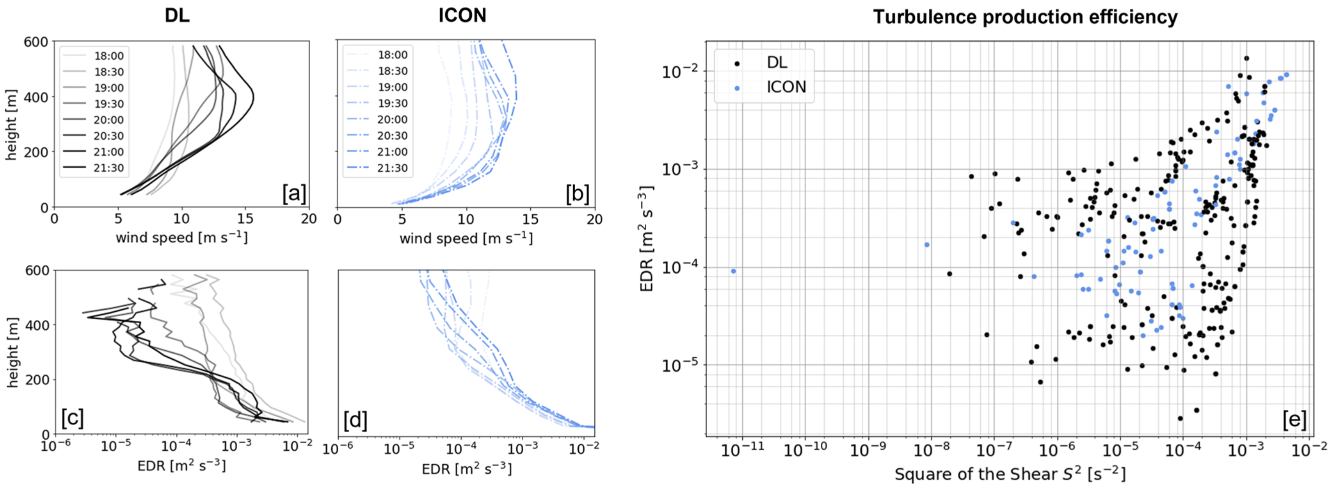

Figure 6. (a–d) A comparison of the temporal evolution of 30-min averaged DL- and ICON-based profiles for wind speed and eddy dissipation rate (EDR) during the transition period from afternoon to evening to the early nocturnal period on 9 June. **(e)** Relationship between vertical wind shear (S^2) and (EDR) to assess turbulence production efficiency.

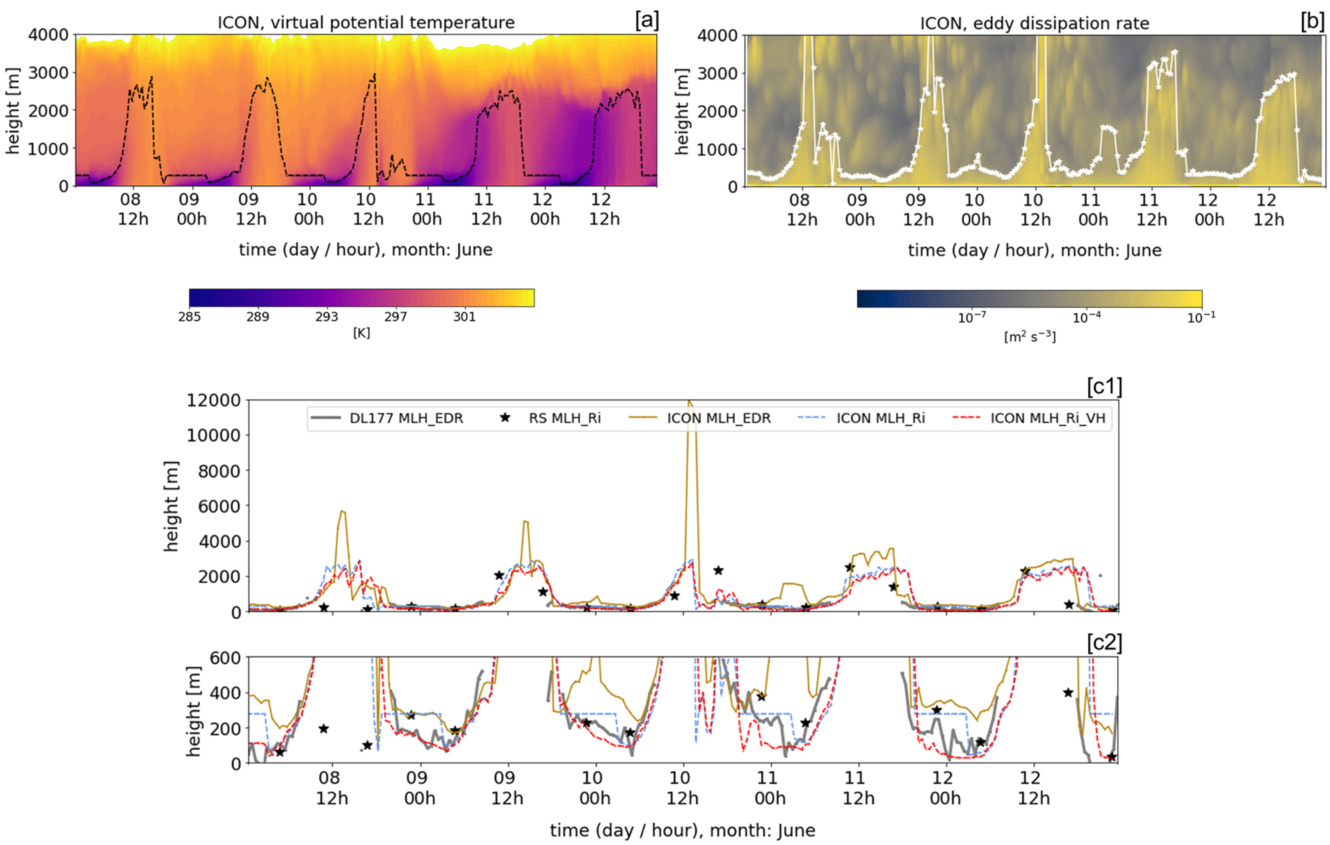

Figure 7. (a) Time-height cross section of simulated virtual potential temperature. The black dashed contour marks the MLH determined using the (offline) MLH diagnostic based on the *Ri* threshold approach with $Ri_{\mathrm{crit}} = 0.2$. **(b)** Simulated time-height cross section of EDR, with a white contour marking the MLH based on the threshold approach using $\mathrm{EDR_{crit}} = 10^{-4}\,\mathrm{m^2\,s^{-3}}$ (see Appendix A). **(c1)** Direct MLH comparison from the different threshold approaches against observations based on both radiosonde soundings and Doppler Lidar derived EDR profiles. **(c2)** Same as in **(c1)** but zoomed into a vertical range of 0–600 m to better resolve nighttime variations.

MLH derived from ICON tends to overestimate the observed MLH. Two aspects contribute to this overestimate. As discussed in Sect. 3 and shown in Figs. 2 and 3, the EDR simulated by ICON is systematically biased high relative to the DL-retrieved EDR. For the purpose of the MLH diagnostic, this could be compensated for by adjusting the EDR threshold used to identify the MLH. Alternatively, the choice of the parameter α_M directly affecting the EDR could be revisited (see Eq. 2). However, this would not address the situation-dependent overmixing (and therefore overestimated EDR) showcased in Sect. 5. Here, the overestimated MLH merely reflects the model's poor representation of the stable boundary layer.

It is a well-known drawback of the Richardson number based diagnostic that the method does not work reliably for stably stratified boundary layers. We see here that the assumed value of 276 m at night in the operational diagnostic would be a reasonable choice, but of course this contains no true information about the state of the mixing layer, and leads to an unrealistic jump in MLH at sunrise. There are alternative diagnostic solutions that combine the *Ri* number based diagnostic with additional criteria for stable conditions (e.g. Liu and Liang, 2010). An update to the ICON post-processing diagnostic is currently under development, and one alternative option considered is the Richardson-number-based diagnostic described in Vogelezang and Holtslag (1996), which differentiates between stable, neutral and unstable conditions and considers a contribution from friction velocity in the shear-production term of the gradient Richardson number calculation (their Eq. 3). The MLH diagnosed thus is included in panel c1–c2 of Fig. 7 as MLH_Ri_VH, and shows generally good agreement with the observed EDR-based MLH. It should be emphasized that the Richardson number method itself is also subject to uncertainty outside the stable ABL. On the first day for example, the RS-MLH diagnostic is unable to reliably detect the top of the daytime convective ABL. On the 10th, the discrepancy between model MLH and RS-MLH in the afternoon can be attributed to the rapid collapse of the ABL in the wake of the simulated cold pool, which was not in fact observed.

The comparisons between EDR- and *Ri*-based MLHs over the last two days also point to a timing mismatch in the representation of boundary layer evolution. There is a clear time shift indicating that the mixing layer grows and decays earlier based on EDR than on *Ri* profiles. Such a behaviour has also been shown in Banakh et al. (2021) and can be explained by the different physical principles and sensitivities of the two methods. EDR responds almost instantaneously to the onset of turbulence, which typically increases rapidly after sunrise due to surface heating. *Ri* adjusts more slowly in response to surface-driven turbulence, causing a delayed detection of MLH growth. Taking these different sensitivities into account, could be a further step towards an improved MLH determination based on NWP data.

Despite the MLH overestimate discussed above, we propose the EDR-based approach as a valuable alternative to complement *Ri*-based formulations under stable conditions, particularly if optimized for systematic model biases in EDR.

6.2 Turbulence intensity diagnostics for wind energy applications

While mean wind speed is a primary driver in wind energy, turbulence intensity (TI) also plays a crucial role by affecting both mechanical loads on wind turbines and their power production. The total TI in wind farms results from ambient atmospheric turbulence and turbine-induced wakes. Although the latter often receives more attention, ambient TI is equally important. It frequently serves as input for power prediction, wake modeling and load calculations, influences wake recovery and is essential for site assessment. Ambient TI is typically obtained from site measurements, empirical models and standard assumptions. This section presents preliminary results that are intended to shed light on the reliability with which the ICON model represents this variable, and on its potential as an additional source of information for ambient TI.

In this study, TI is estimated from the TKE using the formulation

$$\mathrm{TI} = 100\frac{\sqrt{2\mathrm{TKE}}}{V_\mathrm{h}} \quad , \tag{10}$$

in [%], where V_h is the horizontal wind speed. This approach follows Tai et al. (2023) as the turbulence parameterization scheme Turbdiff provides only the SGS-TKE in its entirety as part of the total TKE, but no parameterized SGS standard deviations for individual wind components. These would be necessary to calculate TI according to the standard wind energy definition: $\mathrm{TI} = \sigma_u/U$, where σ_u represents the standard deviation of the longitudinal wind component and U is the streamwise mean wind speed (International Electrotechnical Commission, 2019). As highlighted in Archer (2025), deriving TI from TKE can lead to an overestimation of TI. Due to this limitation, the TI values derived from TKE can only be considered a proxy for IEC-standard TI and the accompanying turbulence conditions. The aim of this section is therefore not to evaluate absolute TI values in accordance with established standards, but to assess whether the quantity derived from the model provides useful information for the application-oriented characterization of ambient turbulence. It is also important to emphasize that, due to data availability, the TI in this analysis is derived from 30 min averages. Wind energy industry standards typically require 10 min averaging periods for TI assessments. Nevertheless, this approach provides a consistent framework for comparing model simulations with turbulence estimates derived from DL, since the DL retrieval method also yields TKE rather than the variances of the individual components.

Figure 8 shows a comparison between the TI value from the ICON simulations and the DL observations, where the TI value is calculated by substituting the respective wind speeds and TKE for both the model and the observation, as shown in Fig. 2, into Eq. (10). Note that TI values above 100 % can occur during strongly non-stationary wind conditions, such as the cold-pool event on june 8th, but they reflect mesoscale variability rather than true turbulence. According to Eq. (10) the peak is driven by two coinciding factors: the very low horizontal wind speed ($v_\mathrm{h} \approx 1\,\mathrm{m\,s^{-1}}$) and a simultaneous peak in TKE, which is likely caused by the high shear and instability at the gust front. Such events are highly relevant for wind energy applications as the simultaneous peak in wind and TI indicates extreme mechanical loads on the turbine structure. Furthermore, the non-stationary nature of cold pool passages poses a challenge for wind power forecasting and turbine control systems. The diurnal evolution of TI is reproduced quite well by the model, however, with a smaller amplitude than observed within the convective ABL (Fig. 8a). Overall the relative model error for TI ranges between $\pm 100\,\%$, but for the majority of the data it is less than $\pm 30\,\%$ (Fig. 8b). This model error is comparable with that for the ICON wind speed simulations (Fig. 2a) but is smaller than the TKE relative error (Fig. 2b). The latter implies that TI, as a normalized turbulence variable, is more robust with respect to systematic errors in wind speed and TKE.

Nonetheless, with $\pm 30\,\%$ the TI bias of the ICON model is relatively large in comparison to the greater precision of more advanced models (e.g. LES) and when compared to accuracy requirements typically associated with detailed wind energy applications. In order to balance accuracy with the amount of effort required, it would be beneficial to determine which practical applications are less sensitive to this TI bias and could therefore benefit from the model's reduced overhead. For instance, in the context of preliminary wind farm site assessment, where the objective is to characterize ambient TI conditions at a broader scale, the model results can still provide valuable insights despite these high levels of uncertainty. A more detailed determination of such acceptance error criteria is, however, beyond the scope of the current work. Here, we deliberately restrict the analysis to a rigorous intercomparison of modeled and observed TI. This provides a brief but transparent and application-independent baseline, which can be used by the wind energy community to judge the suitability of the data for their respective purposes. Future efforts within both the modeling and experimental communities should investigate whether the identified TI model error can be transferred to other locations and whether it is small enough to describe relative TI differences across the entire model domain sufficiently well. ICON wind data are already provided for wind energy applications. Providing the wind energy sector with ICON-based TI products in addition to the existing data would make sense if good agreement can be demonstrated between the TI simulated by ICON and the TI observed from DL measurements across the entire model domain.

Finally, to illustrate the practical implications of the identified discrepancies in wind speed and TI variables from a slightly different perspective, we conducted an additional diagnostic experiment in which both modeled and observed wind speeds and TI were used as input data for a simplified model to estimate wind power production (WPP). It is important to emphasize that this analysis is not intended as a validation of the model results for power prediction applications. Rather, it serves as a sensitivity-type assessment to better understand the relative impact of errors in wind speed and TI on a representative downstream application. In particular, WPP was estimated for a virtual "single turbine park" based on a parametric model of the power curve of wind turbines (Saint-Drenan et al., 2020), for two cases: (i) using wind speed only; and (ii) using wind speed and TI from DL and ICON as input variables. The results are shown in Fig. 9 and are valid for a fictitious wind turbine with a nominal output of 5 MW and a rotor diameter and hub height of 130 m. The results clearly show that for modeled and observed input data, including TI in the wind power estimation leads to a systematic increase in the predicted output for WPP, particularly under convective conditions and at lower wind speeds (Fig. 9a and c). However, since turbine power output depends approximately on the third power of wind speed, the results additionally show that not only large wind speed errors in the simulated nocturnal LLJs but even moderate wind speed biases in the convective boundary layer (Fig. 2) dominate the resulting power errors, masking any potential improvements from those TI information (Fig. 9a and b). This finding suggests that, for this type of application, improving TI estimates alone is unlikely to yield substantial benefits unless accompanied by corresponding improvements in wind speed predictions.

7 Conclusions

This study presents the first application of internally consistent Doppler lidar (DL) retrievals of wind and turbulence profiles – using the method of Smalikho and Banakh (2017) – to evaluate the ICON numerical weather prediction model run with 2.1 km horizontal mesh size, the resolution used at the German Weather Service for regional forecast system over central Europe. Focusing on turbulence kinetic energy (TKE), eddy dissipation rate (EDR) and the turbulent length scale (L_V), which correspond to key equivalent variables in the Turbdiff turbulence scheme, we demonstrate that these DL measurements are well suited for model evaluation and can serve as a complementary evaluation tool alongside LES benchmark simulations. Unlike reference LES simulations, which are often based on simplified and idealized assumptions, DL profile observations reflect more realistic atmospheric conditions and are therefore particularly useful

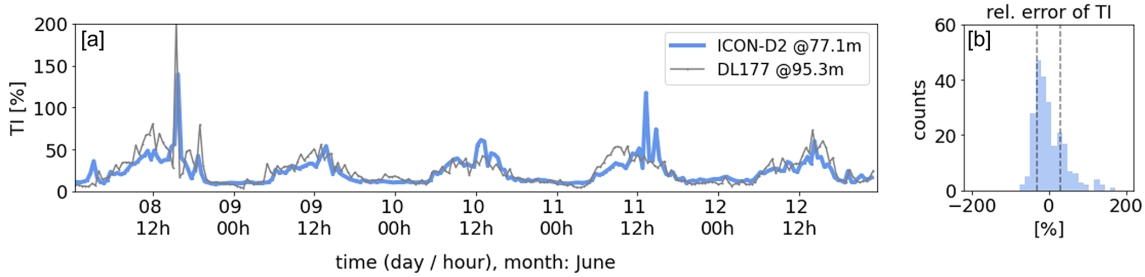

Figure 8. (a) Comparison of diagnosed model TI between ICON simulations at 77 m height with DL-derived TI at 95 m height at the Falkenberg observation site. **(b)** Histogram of relative errors for total TI between ICON and Doppler lidar. Vertical dashed lines indicate a relative error of ±30 %.

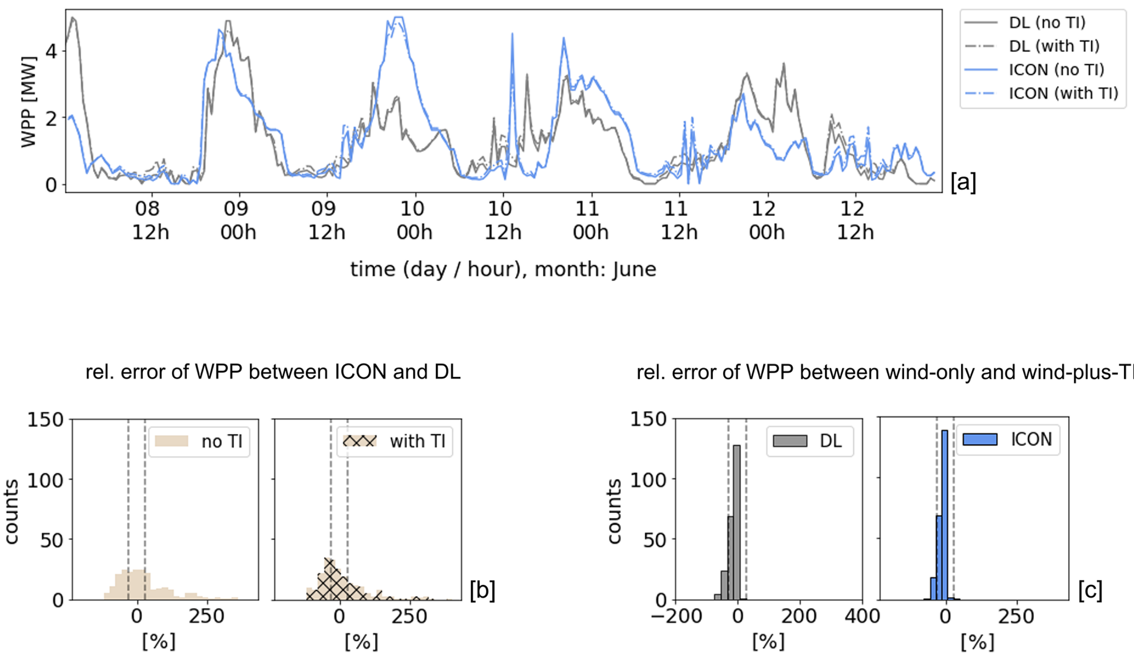

Figure 9. Comparison of Wind Power Production (WPP) based on a parametric model using wind speed and TI data as input variable; **(a)** Estimated WPP from DL observations and ICON output with and without using TI in the calculations **(b)** Relative error of WPP estimates based on ICON data compared to WPP estimates based on DL data once with TI and once without TI as input data. **(c)** Relative error of WPP estimates by neglecting TI once based on DL data and once based on ICON data as input data.

for evaluating turbulence parameterizations under real-world conditions. As the DL measurements were obtained at the Lindenberg Meteorological Observatory in eastern Germany within the lowest 600 m of the atmospheric boundary layer (ABL) over a five-day summer period, the results obtained in this study are specific to this location and the atmospheric conditions encountered, which were characterized by a convective boundary layer during the day and stable near-surface conditions that favor the development of low level jets (LLJs) at night.

The DL scan geometry dictates the spatio-temporal scales of the wind and turbulent properties observed, and both grid-scale (GS) and subgrid-scale (SGS) contributions to the equivalent model properties are accounted for to allow a fair comparison. We show that ICON captures the large scale structures in the lower ABL wind and turbulence profiles and their diurnal evolution reasonably well. Nevertheless, more detailed analyses of time series at individual heights and composite profiles at different times of day reveal substantial smaller scale discrepancies. The largest ones are found in the representation of cold pools and LLJs, affecting both the wind field and the associated turbulence characteristics. While exact agreement on the timing, intensity and location of cold pools cannot be expected, the model errors associated with LLJs are more systematic. Overall, wind errors are smaller than those in the turbulence-related quantities, which is expected given the additional challenges associated with turbulence parameterization. Among the turbulence variables the model-observation discrepancies are larger for EDR than for TKE. The minimum diffusivity prescribed in Turbdiff may offer one possible explanation for this. Although diffusivity and dissipation are not directly equivalent, stronger mixing can promote turbulence regimes with increased dissipation rates, thereby contributing to the larger EDR bias. Sandu et al. (2013) describe how artificially enhanced turbulent diffusivity in the NWP model of the European Centre for Medium-Range Weather Forecasts compensates for biases in other processes (such as orographic drag and land-atmosphere coupling). This improves forecast scores while leading to a poorer representation of winds in stably stratified boundary layers. The similarities in the observed model behavior suggest that some of these conclusions may also apply to ICON.

Model performance also varies with time of day. Agreement is better during daytime and the evening transition period, corresponding mainly to unstable atmospheric conditions and the transition from weakly unstable to weakly stable ABL regimes. Larger discrepancies occur at night and during morning transitions, i.e. when atmospheric conditions are mainly stable. While daytime conditions are characterized by a slightly underestimated TKE and overestimated EDR, nighttime conditions show an overestimation in both quantities. This indicates that the balance between turbulence production and dissipation is represented differently under convective and stable ABL conditions, i.e. turbulence is dissipated too rapidly under convective conditions, while excessive turbulent mixing persist under stable conditions.

The relative contributions from GS and SGS to the total TKE give some insights into the model's ability to flexibly repartition TKE according to the prevailing scales of the turbulent circulations. As expected for the grid-size resolution of the ICON D2 configuration with 2.1 km, SGS-TKE generally dominates over GS-TKE. There is a pronounced diurnal cycle, reflecting the dynamic partitioning of turbulent energy between resolved and unresolved scales over the course of the day. Enhanced GS contributions are primarily associated with cold-pool events and LLJs. The model thus demonstrates a degree of scale-awareness yielding an overall reasonable total TKE without indications for obvious problems, such as double-counting. Future work is planned to investigate the model's scale adaptivity further by evaluating turbulent properties across multiple sub-kilometer scale grid resolutions. The coincidence of larger model errors in wind and turbulence variables and enhanced GS-TKE contributions during cold pool and LLJ events could be an indication that the representation of turbulence becomes more challenging in gray-zone situations, where energy is increasingly partitioned toward partially resolved turbulence structures.

The turbulent length scale L_V was further used as indicator of how effectively the model resolves turbulent scales. The DL retrieval of this parameter shows a clear diurnal cycle which is much weaker when the DL-equivalent integral length scale L_V^* is derived from model data. While some of this discrepancy may be attributed to an imperfect accounting of the model TKE contained in unresolved, organized circulations which are not considered part of the subgrid-range covered by the turbulence parameterization, the generally good agreement of the total TKE means that this argument is unlikely to explain the differences in their entirety. We can demonstrate that the behavior of L_V^* is directly related to the definition of the integral turbulent length scale l_B representing the upper limit of the SGS turbulence spectrum parameterized by Turbdiff. l_B solely depends on mesh size and distance to the surface, which are flow-invariant. A possible step towards a better model performance in this context might be achieved by adopting a stability-dependent length scale formulation.

Over the five-day period, the second night stands out with excessive ABL mixing contributing to momentum fluxes which distort the correct evolution of LLJs compared to measured DL wind profiles. While the observations show that low-turbulence states can be maintained under strong shear conditions, in the model shear-induced turbulence is not adequately suppressed. This behavior is consistent with the weak sensitivity to stability observed in the turbulent length-scale analysis but may also point to deficiencies in the formulation of the stability function within Turbdiff. Two alternative turbulence parameterizations available for ICON that might be of particular interest are the TKE scheme with scalar variance closure (TKE-SV) developed by Machulskaya and Mironov (2013, 2020), and the two-energies turbulence (2TE) scheme by Bašták Ďurán et al. (2018, 2022). Both of these include more sophisticated treatment of the stability functions and/or turbulent length scale formulations.

Beyond the model evaluation itself, the final analyses addresses the potential value of the ICON wind and turbulence simulations for practical forecasting applications. One application investigated in this study is the retrieval of the mixing layer height (MLH) from EDR profiles for example for dispersion modeling. Following common practice for DL observations, a threshold-based retrieval approach was used and transferred to the simulated EDR profiles. ICON currently uses a MLH diagnostic derived from Richardson number profiles. In the absence of reliable MLH estimates from this method, a constant value is assumed for the nocturnal MLH. We demonstrate that a MLH diagnostic based on the EDR threshold method could provide a more physically realistic alternative particularly for the stable boundary layer and rapidly changing growth and decay phases of the convective ABL, provided that the systematic EDR bias in ICON is considered.

Another application analyzed was the model's performance in providing ambient "turbulence intensity" (TI; a parameter combining wind and TKE information) for wind energy applications. While ambient TI should not be interpreted as the total turbulence intensity experienced by wind turbines, it nevertheless represents an important environmental boundary condition influencing the generation and evolution of turbine wakes. In ICON, the modeled ambient TI is more robust than TKE alone with model errors largely ranging around $\pm 30\,\%$. Although TI could theoretically provide additional benefit for wind power estimates, particularly during convective ABL conditions, we find that the combined errors in the simulated wind field and TI are currently too large to reliably exploit this potential. Even the errors identified in LLJs during stable conditions at night led to significant uncertainty in the WPP estimates, regardless of the TI effect. The results highlight the importance of further improving the ICON simulation of wind and turbulence variables to provide better support for future wind energy applications. Until then, further research in collaboration with wind energy experts is needed to assess whether the simulated TI can provide useful information in areas other than wind power forecasting, such as estimating the load on wind turbines, despite the identified model error in the prediction of ambient TI.

By approaching the evaluation of Turbdiff from different perspectives we have thus demonstrated the utility and versatility of the Doppler lidar observations to evaluate various aspects of the model winds and turbulent properties, yielding useful insights for future model development and forecast applications. We point out that efforts to refine model parameterizations could benefit from advanced observations such as DL measurements to support a consistent improvement of the overall flow representation. This perspective suggests that DL observations, in addition to their role in evaluation, may offer a helpful avenue for the development and tuning of model parameterizations. One drawback of the evaluation method discussed in this work is that it is limited to the lowest 600 m of the ABL where a concurrent retrieval of winds and turbulent properties from the DL are possible. Given that the greatest remaining model biases in ICON occur at night under stable conditions, this detracts only slightly from the utility of the method. Our study here is limited to a single location, but the network of DLs in Germany and across Europe (ACTRIS; Laj et al., 2024) is steadily growing, offering the potential for a wider application. The record of DL observations at Lindenberg now spans several years, covering a large variety of different ABL cases for further study.

Appendix A: Evaluation of DL-derived MLH estimated from EDR profiles

Figure A1 shows a comparison of different MLH estimates derived from DL profiles for EDR, and from vertical profiles of wind and temperature obtained from radiosonde soundings (hereafter RS). The MLH derived from DL-based EDR profiles (hereafter DL_MLH_EDR) is obtained by seeking for that height at which the EDR falls below a threshold of $\mathrm{EDR_{crit}} = 10^{-4}\,\mathrm{m^2\,s^{-3}}$. This is in analogy with the approach described in O'Connor et al. (2010) and Banakh et al. (2021). The various MLHs derived from RS are the result of applying different definitions described in Beyrich and Leps (2012) which ultimately result in different MLH results. The criteria used here to determine the MLH include for instance the determination of temperature inversions, the identification of zones with significant wind shear and the analysis of Richardson number (Ri) profiles. The corresponding derived MLHs are referred to as RS_MLH_DTDH, RS_MLH_DSDH, and RS_MLH_Ri in the following. However, these represent only a subset among the variety of existing approaches suggested in Beyrich and Leps (2012). The comparisons shown in Fig. A1a indicate that at night and during the morning transitions DL_MLH_EDR is always in good agreement with at least one of the RS based MLHs from the various methods. Differences in the derived MLHs are mostly < 200 m. This provides evidence for reliable MLH observations from DL-derived EDR profiles even under stable conditions with strong wind shear. Under these conditions, shear-induced TKE and associated increased dissipation rates can also occur outside the actual mixing layer, which theoretically could have complicated a correct determination of the MLH. During the evening transition the spread among the RS based MLHs from the various methods can be quite large. In such cases best agreement occurs between DL_MLH_EDR and RS_MLH_DTDH. Note that DL_MLH_EDR values above 600 m, i.e. for convective boundary layers with a depth exceeding this value cannot be derived from the present data set. This is a consequence of the scanning and retrieval method used for the DL measurements. For MLH obtained from DL-based EDR profiles throughout the whole day, the scanning procedure would have to be adapted according to the approach described in Banakh et al. (2021). Overall, the comparisons show the consistency and reliability of DL-based MLH estimates relative to established radiosonde-based approaches. Nevertheless, exact agreement is not anticipated because the applied retrieval methods are based on different definitions and physical characteristics of the mixing layer height.

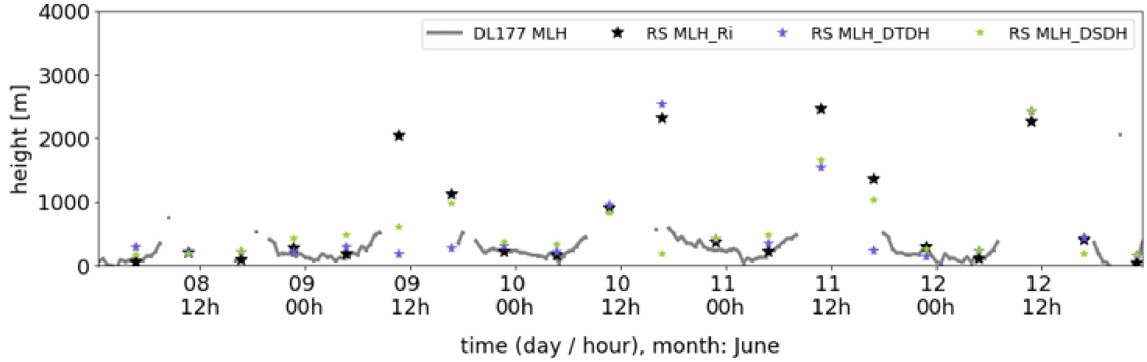

Figure A1. (a) Comparison of the mixing layer height (MLH) obtained from Doppler lidar (DL) and radiosonde soundings (RS) for the 5 d time period 8–12 June 2023 at the MOL-RAO site. DL177 MLH denotes the MLH obtained from DL based EDR profiles. RS_MLH_DTDH, RS_MLH_DSDH, and RS_MLH_Ri denote different RS based MLHs according to specific definitions (e.g. existence of temperature inversion, zones with significant wind shear or the height at which the Richardson number exceeds a critical level) applied to RS based temperature and wind profiles.

Code and data availability. The source code for ICON is available at https://www.icon-model.org/ (last access: 2 May 2026) via gitlab repository under a permissive open source licence (BSD-3C). For this study, release tag icon-2026.04 was used, and is archived as a static copy together with the Doppler lider retrieval, the model grids, external parameters and forcing data to recreate the simulations on Zenodo (https://doi.org/10.5281/zenodo.20161993, Ahlgrimm and Päschke, 2025).

Author contributions. Both authors have contributed equally to this work. MA provided the research objective while the broader conceptual framework was developed collaboratively. MA ran the model simulations and introduced additional diagnostics to produce the model-derived wind and turbulence products. EP produced the Doppler Lidar retrieval, developed the methodology and conducted the formal analysis. Both authors have contributed to the visualisation and writing of this manuscript.

Competing interests. The contact author has declared that neither of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. Special thanks our DWD colleagues Ekaterina Machulskaya, Dmitrii Mironov, Matthias Raschendorfer and Günther Zängl for helpful advice on ICON's turbulence schemes and to Frank Beyrich from the Lindenberg Observatory for fruitful discussion around Doppler Lidar usage.

Financial support. This research was supported by the German Weather Service (Deutscher Wetterdienst) and the Hans-Ertel-Centre for Weather Research (Hans-Ertel-Zentrum für Wetterforschung), funded by the Federal Ministry of Transport (Bundesministerium für Verkehr; BMV).

Review statement. This paper was edited by Nicola Bodini and reviewed by Brigitta Goger and two anonymous referees.

References

Ahlgrimm, M. and Päschke, E.: Model code, forcing data and Doppler lidar retrieval, Zenodo [code] https://doi.org/10.5281/zenodo.20161993, 2025.

Archer, C. L.: Brief communication: A note on the variance of wind speed and turbulence intensity, Wind Energ. Sci., 10, 1433–1438, https://doi.org/10.5194/wes-10-1433-2025, 2025.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities, Mon. Weather Rev., 139, 3887–3905, https://doi.org/10.1175/MWR-D-10-05013.1, 2011.

Banakh, V. and Smalikho, I.: Coherent Doppler Wind Lidars in a Turbulent Atmosphere, Artech House Publishers, Boston, USA, illustrated edn., ISBN 9781608076680, 2013.

Banakh, V. A. and Smalikho, I. N.: Lidar estimates of the anisotropy of wind turbulence in a stable atmospheric boundary layer, Remote Sens.-Basel, 11, https://doi.org/10.3390/rs11182115, 2019.

Banakh, V. A., Smalikho, I. N., and Falits, A. V.: Estimation of the height of the turbulent mixing layer from data of Doppler lidar measurements using conical scanning by a probe beam, Atmos. Meas. Tech., 14, 1511–1524, https://doi.org/10.5194/amt-14-1511-2021, 2021.

Bašták Ďurán, I., Geleyn, J.-F., Váňa, F., Schmidli, J., and Brožková, R.: A turbulence scheme with two prognostic turbulence energies, J. Atmos. Sci., 75, 3381–3402, https://doi.org/10.1175/JAS-D-18-0026.1, 2018.

Bašták Ďurán, I., Sakradzija, M., and Schmidli, J.: The two-energies turbulence scheme coupled to the assumed PDF method, J. Adv. Model. Earth Sy., 14, e2021MS002922, https://doi.org/10.1029/2021MS002922, 2022.

Beyrich, F. and Leps, J.-P.: An operational mixing height data set from routine radiosoundings at Lindenberg: methodology, Meteor. Z., 21, 337–348, https://doi.org/10.1127/0941-2948/2012/0333, 2012.

Blackadar, A. K.: The vertical distribution of wind and turbulent exchange in a neutral atmosphere, J. Geophys. Res., 67, 3095–3102, https://doi.org/10.1029/JZ067i008p03095, 1962.

Boutle, I. A., Finnenkoetter, A., Lock, A. P., and Wells, H.: The London Model: forecasting fog at 333 m resolution, Q. J. Roy. Meteor. Soc., 142, 360–371, https://doi.org/10.1002/qj.2656, 2016.

Buzzi, M., Rotach, M. W., Holtslag, M., and Holtslag, A. A.: Evaluation of the COSMO-SC turbulence scheme in a shear-driven stable boundary layer, Meteor. Z., 20, 335, https://doi.org/10.1127/0941-2948/2011/0050, 2011.

Dewani, N., Sakradzija, M., Schlemmer, L., Leinweber, R., and Schmidli, J.: Dependency of vertical velocity variance on meteorological conditions in the convective boundary layer, Atmos. Chem. Phys., 23, 4045–4058, https://doi.org/10.5194/acp-23-4045-2023, 2023.

Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin, R., Schulz, J.-P., and Vogel, G.: A Description of the Non-hydrostatic Regional COSMO-Model, Part II, Physical Parameterizations, https://doi.org/10.5676/DWD_pub/nwv/cosmo-doc_6.00_II, 2021.

Doubrawa, P. and Muñoz-Esparza, D.: Simulating real atmospheric boundary layers at gray-zone resolutions: how do currently available turbulence parameterizations perform?, Atmosphere-Basel, 11, 345, https://doi.org/10.3390/atmos11040345, 2020.

Dudhia, J.: Challenges in sub-kilometer grid modeling of the convective planetary boundary layer, Meteorology, 1, 402–413, https://doi.org/10.3390/meteorology1040026, 2022.

Finn, T. S., Geppert, G., and Ament, F.: Towards assimilation of wind profile observations in the atmospheric boundary layer with a sub-kilometrescale ensemble data assimilation system, Tellus B, https://doi.org/10.1080/16000870.2020.1764307, 2020.

Glazer, R. H., Grimmond, S., Blunn, L., Fenner, D., Lean, H., Christen, A., Morrison, W., Looschelders, D., and Shonk, J. K. P.: Hectometric-scale modelling of the mixed layer in an urban region evaluated with a dense LiDAR-ceilometer network, Weather Clim. Dynam., 6, 1723–1742, https://doi.org/10.5194/wcd-6-1723-2025, 2025.

Goecke, T. and Machulskaya, E.: Aviation turbulence forecasting at DWD with ICON: Methodology, case studies, and verification, Mon. Weather Rev., 149, 2115–2130, https://doi.org/10.1175/MWR-D-19-0383.1, 2021.

Goger, B. and Dipankar, A.: The impact of mesh size, turbulence parameterization, and land-surface-exchange scheme on simulations of the mountain boundary layer in the hectometric range, Q. J. Roy. Meteor. Soc., 150, 3853–3873, https://doi.org/10.1002/qj.4799, 2024.

Goger, B., Rotach, M. W., Gohm, A., Fuhrer, O., Stiperski, I., and Holtslag, A. A.: The impact of three-dimensional effects on the simulation of turbulence kinetic energy in a major alpine valley, Bound.-Lay. Meteorol., 168, 1–27, 2018.

Harvey, N. J., Hogan, R. J., and Dacre, H. F.: A method to diagnose boundary-layer type using Doppler lidar, Q. J. Roy. Meteor. Soc., 139, 1681–1693, https://doi.org/10.1002/qj.2068, 2013.

He, Y., McFarlane, N. A., and Monahan, A. H.: A New TKE-based parameterization of atmospheric turbulence in the Canadian global and regional climate models, J. Adv. Model. Earth Sy., 11, 1153–1188, https://doi.org/10.1029/2018MS001532, 2019.

Hohenegger, C., Ament, F., Beyrich, F., Löhnert, U., Rust, H., Bange, J., Böck, T., Böttcher, C., Boventer, J., Burgemeister, F., Clemens, M., Detring, C., Detring, I., Dewani, N., Duran, I. B., Fiedler, S., Göber, M., van Heerwaarden, C., Heusinkveld, B., Kirsch, B., Klocke, D., Knist, C., Lange, I., Lauermann, F., Lehmann, V., Lehmke, J., Leinweber, R., Lundgren, K., Masbou, M., Mauder, M., Mol, W., Nevermann, H., Nomokonova, T., Päschke, E., Platis, A., Reichardt, J., Rochette, L., Sakradzija, M., Schlemmer, L., Schmidli, J., Shokri, N., Sobottke, V., Speidel, J., Steinheuer, J., Turner, D. D., Vogelmann, H., Wedemeyer, C., Weide-Luiz, E., Wiesner, S., Wildmann, N., Wolz, K., and Wetz, T.: FESSTVaL: the field experiment on submesoscale spatio-temporal variability in Lindenberg, B. Am. Meteorol. Soc., 104, E1875 – E1892, https://doi.org/10.1175/BAMS-D-21-0330.1, 2023.

Honnert, R., Masson, V., and Couvreux, F.: A diagnostic for evaluating the representation of turbulence in atmospheric models at the kilometric scale, J. Atmos. Sci., 68, 3112–3131, https://doi.org/10.1175/JAS-D-11-061.1, 2011.

Honnert, R., Efstathiou, G. A., Beare, R. J., Ito, J., Lock, A., Neggers, R., Plant, R. S., Shin, H. H., Tomassini, L., and Zhou, B.: The atmospheric boundary layer and the "gray zone" of turbulence: a critical review, J. Geophys. Res.-Atmos., 125, e2019JD030317, https://doi.org/10.1029/2019JD030317, 2020.

International Electrotechnical Commission: Wind energy generation systems – Part 1: Design requirements, Technical Report IEC 61400-1 Ed. 4.0 B:2019, IEC, Geneva, Switzerland, https://webstore.ansi.org/standards/iec/iec61400ed2019-2419167?source=blog (last access: 23 April 2025), 2019.

Joe, P., Belair, S., Bernier, N. B., Bouchet, V., Brook, J. R., Brunet, D., Burrows, W., Charland, J.-P., Dehghan, A., Driedger, N., Duhaime, C., Evans, G., Filion, A.-B., Frenette, R., de Grandpré, J., Gultepe, I., Henderson, D., Herdt, A., Hilker, N., Huang, L., Hung, E., Isaac, G., Jeong, C.-H., Johnston, D., Klaassen, J., Leroyer, S., Lin, H., MacDonald, M., MacPhee, J., Mariani, Z., Munoz, T., Reid, J., Robichaud, A., Rochon, Y., Shairsingh, K., Sills, D., Spacek, L., Stroud, C., Su, Y., Taylor, N., Vanos, J., Voogt, J., Wang, J. M., Wiechers, T., Wren, S., Yang, H., and Yip, T.: The environment Canada pan and parapan American science showcase project, B. Am. Meteorol. Soc., 99, 921–953, https://doi.org/10.1175/BAMS-D-16-0162.1, 2018.

Kirsch, B., Ament, F., and Hohenegger, C.: Convective cold pools in long-term boundary layer mast observations, Mon. Weather Rev., 149, 811–820, https://doi.org/10.1175/MWR-D-20-0197.1, 2021.

Kirsch, B., Hohenegger, C., and Ament, F.: Morphology and growth of convective cold pools observed by a dense station network in Germany, Q. J. Roy. Meteor. Soc., 150, 857–876, https://doi.org/10.1002/qj.4626, 2024.

Kolmogorov, A. N.: A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number, J. Fluid Mech., 13, 82–85, https://doi.org/10.1017/S0022112062000518, 1962.

Kropfli, R. A.: Single Doppler radar measurements of turbulence profiles in the convective boundary layer, J. Atmos. Ocean. Tech., 3, 305–314, https://doi.org/10.1175/1520-0426(1986)003%3C0305:SDRMOT%3E2.0.CO;2, 1986.

Laj, P., Myhre, C. L., Riffault, V., Amiridis, V., Fuchs, H., Eleftheriadis, K., Petäjä, T., Salameh, T., Kivekäs, N., Juurola, E., Saponaro, G., Philippin, S., Cornacchia, C., Alados Arboledas, L., Baars, H., Claude, A., De Mazière, M., Dils, B., Dufresne, M., Evangeliou, N., Favez, O., Fiebig, M., Haeffelin, M., Herrmann, H., Höhler, K., Illmann, N., Kreuter, A., Ludewig, E., Marinou, E., Möhler, O., Mona, L., Murberg, L. E., Nicolae, D., Novelli, A., O'Connor, E., Ohneiser, K., Petracca Altieri, R. M., Picquet-Varrault, B., van Pinxteren, D., Pospichal, B., Putaud, J.-P., Reimann, S., Siomos, N., Stachlewska, I., Tillmann, R., Voudouri, K. A., Wandinger, U., Wiedensohler, A., Apituley, A., Comerón, A., Gysel-Beer, M., Mihalopoulos, N., Nikolova, N., Pietruczuk, A., Sauvage, S., Sciare, J., Skov, H., Svendby, T., Swietlicki, E., Tonev, D., Vaughan, G., Zdimal, V., Baltensperger, U., Doussin, J.-F., Kulmala, M., Pappalardo, G., Sorvari Sundet, S., and Vana, M.: Aerosol, clouds and trace gases research infrastructure (ACTRIS): the European research infrastructure supporting atmospheric science, B. Am. Meteorol. Soc., 105, E1098–E1136, https://doi.org/10.1175/BAMS-D-23-0064.1, 2024.

Liu, S. and Liang, X.-Z.: Observed diurnal cycle climatology of planetary boundary layer height, J. Climate, 23, 5790–5809, https://doi.org/10.1175/2010JCLI3552.1, 2010.

Lothon, M., Lenschow, D. H., and Mayor, S. D.: Doppler lidar measurements of vertical velocity spectra in the convective planetary boundary layer, Bound.-Lay. Meteorol., 132, 205–226, https://doi.org/10.1007/s10546-009-9398-y, 2009.

Machulskaya, E. and Mironov, D.: Implementation of TKE-scalar variance mixing scheme into COSMO, COSMO Newsletter, 25–33, https://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter13/cnl13_03.pdf (last access: 2 May 2026), 2013.

Machulskaya, E. and Mironov, D. V.: The stability functions and realizability of the turbulent kinetic energy–scalar variance closure for moist atmospheric boundary-layer turbulence, Bound.-Lay. Meteorol., 176, 197–228, https://doi.org/10.1007/s10546-020-00528-7, 2020.

Manninen, A., Marke, T., Tuononen, M., and O'Connor, E.: Atmospheric boundary layer classification with Doppler lidar, J. Geophys. Res.-Atmos., 123, 8172–8189, https://doi.org/10.1029/2017JD028169, 2018.

Mellor, G. L. and Yamada, T.: Development of a turbulence closure model for geophysical fluid problems, Rev. Geophys., 20, 851–875, https://doi.org/10.1029/RG020i004p00851, 1982.

Mora, D. and Obligado, M.: Estimating the integral length scale on turbulent flows from the zero crossings of the longitudinal velocity fluctuation, Exp. Fluids, 61, 199, https://doi.org/10.1007/s00348-020-03033-2, 2020.

Muñoz-Esparza, D., Sharman, R. D., and Lundquist, J. K.: Turbulence dissipation rate in the atmospheric boundary layer: observations and WRF mesoscale modeling during the XPIA field campaign, Mon. Weather Rev., 146, 351–371, https://doi.org/10.1175/MWR-D-17-0186.1, 2018.

Nilsson, E., Lohou, F., Lothon, M., Pardyjak, E., Mahrt, L., and Darbieu, C.: Turbulence kinetic energy budget during the afternoon transition – Part 1: Observed surface TKE budget and boundary layer description for 10 intensive observation period days, Atmos. Chem. Phys., 16, 8849–8872, https://doi.org/10.5194/acp-16-8849-2016, 2016a.

Nilsson, E., Lothon, M., Lohou, F., Pardyjak, E., Hartogensis, O., and Darbieu, C.: Turbulence kinetic energy budget during the afternoon transition – Part 2: A simple TKE model, Atmos. Chem. Phys., 16, 8873–8898, https://doi.org/10.5194/acp-16-8873-2016, 2016b.

O'Connor, E. J., Illingworth, A. J., Brooks, I. M., Westbrook, C. D., Hogan, R. J., Davies, F., and Brooks, B. J.: A method for estimating the turbulent kinetic energy dissipation rate from a vertically pointing Doppler lidar, and independent evaluation from balloon-borne in situ measurements, J. Atmos. Ocean. Tech., 27, 1652–1664, https://doi.org/10.1175/2010JTECHA1455.1, 2010.

Päschke, E. and Detring, C.: Noise filtering options for conically scanning Doppler lidar measurements with low pulse accumulation, Atmos. Meas. Tech., 17, 3187–3217, https://doi.org/10.5194/amt-17-3187-2024, 2024.

Raschendorfer, M.: The new turbulence parametrization of LM, COSMO Newsletter, 1, 89–97, http://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter01/newsLetter_01.pdf (last access: 2 May 2026), 2001.

Raschendorfer, M., Simmer, C., and Gross, P.: Parameterization of Turbulent Transport in the Atmosphere, Vol. 97, Springer, https://doi.org/10.1007/3-540-45256-7_10, 2003.

Richardson, L. F.: Weather Prediction by Numerical Process, University Press, https://doi.org/10.1017/CBO9780511618291, 1922.

Rohanizadegan, M., Petrone, R. M., Pomeroy, J. W., and Kosovic, B.: Analysis of turbulence and turbulence kinetic energy dynamics in complex terrain, J. Geophys. Res.-Atmos., 130, e2023JD040558, https://doi.org/10.1029/2023JD040558, 2025.

Rotta, J.: Statistische Theorie nichthomogener Turbulenz, Z. Phys., 129, 547–572, https://doi.org/10.1007/BF01330059, 1951.

Saint-Drenan, Y.-M., Besseau, R., Jansen, M., Staffell, I., Troccoli, A., Dubus, L., Schmidt, J., Gruber, K., Simões, S. G., and Heier, S.: A parametric model for wind turbine power curves incorporating environmental conditions, Renew. Energ., 157, 754–768, https://doi.org/10.1016/j.renene.2020.04.123, 2020.

Sakradzija, M., Ahlgrimm, M., Beyrich, F., van Heerwaarden, C., Päschke, E., Görsdorf, U., Kirsch, B., Schmidli, J., and Dewani, N.: Where numerical weather prediction meets large-eddy simulations in a convective boundary layer during the

FESSTVaL field experiment, Q. J. Roy. Meteor. Soc., e70037, https://doi.org/10.1002/qj.70037, 2025.

Sandu, I., Beljaars, A., Bechtold, P., Mauritsen, T., and Balsamo, G.: Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models?, J. Adv. Model. Earth Sy., 5, 117–133, https://doi.org/10.1002/jame.20013, 2013.

Shin, H. H. and Hong, S.-Y.: Analysis of resolved and parameterized vertical transports in convective boundary layers at gray-zone resolutions, J. Atmos. Sci., 70, 3248–3261, https://doi.org/10.1175/JAS-D-12-0290.1, 2013.

Smalikho, I. N. and Banakh, V. A.: Measurements of wind turbulence parameters by a conically scanning coherent Doppler lidar in the atmospheric boundary layer, Atmos. Meas. Tech., 10, 4191–4208, https://doi.org/10.5194/amt-10-4191-2017, 2017.

Sørensen, J., Rasmussen, A., and Svensmark, H.: Forecast of atmospheric boundary-layer height utilised for ETEX real-time dispersion modelling, Phys. Chem. Earth, 21, 435–439, https://doi.org/10.1016/S0079-1946(97)81138-X, 1996.

Tai, S.-L., Berg, L. K., Krishnamurthy, R., Newsom, R., and Kirincich, A.: Validation of turbulence intensity as simulated by the Weather Research and Forecasting model off the US northeast coast, Wind Energ. Sci., 8, 433–448, https://doi.org/10.5194/wes-8-433-2023, 2023.

Vogelezang, D. and Holtslag, A.: Evaluation and model impacts of alternative boundary-layer height formulations, Bound.-Lay. Meteorol., 81, 245–269, https://doi.org/10.1007/BF02430331, 1996.

Welford, B. P.: Note on a method for calculating corrected sums of squares and products, Technometrics, 4, 419–420, https://doi.org/10.1080/00401706.1962.10490022, 1962.

Wildmann, N., Päschke, E., Roiger, A., and Mallaun, C.: Towards improved turbulence estimation with Doppler wind lidar velocity-azimuth display (VAD) scans, Atmos. Meas. Tech., 13, 4141–4158, https://doi.org/10.5194/amt-13-4141-2020, 2020.

Wyngaard, J. C.: Toward numerical modeling in the "Terra Incognita", J. Atmos. Sci., 61, 1816–1826, https://doi.org/10.1175/1520-0469(2004)061<1816:TNMITT>2.0.CO;2, 2004.

Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M.: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: description of the non-hydrostatic dynamical core, Q. J. Roy. Meteor. Soc., 141, 563–579, https://doi.org/10.1002/qj.2378, 2015.

Zonato, A., Martilli, A., Jimenez, P. A., Dudhia, J., Zardi, D., and Giovannini, L.: A new K–ε turbulence parameterization for mesoscale meteorological models, Mon. Weather Rev., 150, 2157–2174, https://doi.org/10.1175/MWR-D-21-0299.1, 2022.