Wind and turbulence evaluation of the ICON model (icon-2024.01-1icon-2026.04) at sub-kilometer scales 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)

Abstract. Regional Turbulence parameterization in numerical weather prediction models are increasingly run at sub-kilometer scale horizontal resolutions, where turbulence can no longer be considered as an entirely sub-grid scale process (the "turbulent gray zone")(NWP) models remains a challenge, particularly as resolution continues to increase. Existing turbulence evaluation methods often rely on high-resolution benchmark simulations . We present an which commonly depend on idealized assumptions and boundary conditions. An alternative evaluation method is presented here, based on Doppler lidar retrievals of winds (DL) retrievals of wind and turbulent properties in the lower atmospheric boundary layer .

Two configurations of the ICOsahedral Nonhydrostatic (ICON) model are compared for horizontal mesh sizes ranging from 2.1 km to 78 m: the operationally used 1D TKE scheme "Turbdiff" which includes scale-adaptive features and accounts for some 3D turbulent terms, and a 3D Smagorinsky scheme, commonly used in large eddy simulations. Winds, 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 diffusivity dissipation rate (EDR) and turbulent length scalefrom the Doppler lidar retrieval are used in the evaluation , 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 for NWP at the German Weather Service. Diagnostics for an equitable comparison of observed and simulated TKE are applied, accounting for grid-scale and sub-grid scale model contributions model contributions on scales equivalent to those of the DL retrieval.

Results show that winds remain very similar across all resolutions and between configurations, while pronounced differences are found in TKE and EDR. The 3D Smagorinsky configuration performs poorly at coarse resolutions, overestimating both 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, but shows comparable performance to Turbdiff at 78m. Turbdiff more successfully re-partitions TKE between grid-scale and subgrid-scale contributions, roughly preserving the total TKE across mesh sizes. Both configurations show systematic errors in the stable nocturnal boundary layer, coinciding with poor representations of the . 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.

This study demonstrates the value of the Doppler lidar retrieval as an evaluation tool, shows that the scale-adaptive Turbdiff scheme suffers only minor loss of performance into the turbulent gray zone and highlights specific aspects of the scheme that limit performance for stable night-time conditions. We illustrate 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 is entirely subgrid-scale, become less applicable at these scales (Honnert et al., 2020; Doubrawa and Muñoz-Esparza, 2020). The appro-priate (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).

The ICOsahedral Nonhydrostatic (ICON) model (Zängl et al., 2015) used for NWP at the German Weather Service (DWD) currently runs with horizontal mesh sizes of 13 km for global ICON simulations, 6.5 km for the European nested region and 2.1 km for the limited area configuration over Germany. The horizontal mesh size is defined as the square root of the triangular grid cell area. A recent milestone is the development and operational launch of the first sub-kilometer application with a mesh size of 500 m over Germany (Reinert et al., 2025). Before its adoption, this latest model configuration was validated with the suite of conventional SYNOP, radiosonde and aircraft observations used routinely at DWD. Benefits of the higher resolution were found primarily over the Alps and German central mountainous areas where a better resolved orography leads to improved 2 m temperature, relative humidity and 10 m winds. However, no process-specific evaluation of the standard ICON turbulence scheme was performed to assess its performance in the gray zone.

Appropriate methods are requiredto evaluate and further develop the turbulence parameterizations used in NWP models for scales approaching the turbulent gray zoneTo guide the further development of turbulence parameterizations, appropriate methods for model evaluation are required. Having suitable reference data is a fundamental requirement 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), and in the operational setting, routine validation against SYNOP observations. 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).

Evaluating models in the gray zone requires methods to account for the separate contribution of grid-scale (hereafter GS) and subgrid-scale (hereafter SGS) turbulence contributions. When evaluating against LES reference cases, the LES data are typically filtered to match the model's effective resolution, ensuring a consistent scale separation (Honnert et al., 2011). This is generally not possible when using measurement data as a reference. Unlike LES data, measurements are typically limited in spatial coverage and resolution, providing only partial information about the turbulent flow to start with. Therefore it is generally more advantageous to compare to the total turbulent quantities measured. This approach then requires the extraction and combination of GS and SGS turbulence contributions from the model for a fair comparison. This may require the implementation of additional diagnostic variables during the simulation, since standard model output typically provides only accumulated or instantaneous fields which do not capture the resolved fluctuations needed for TKE or variance validation (Tai et al., 2023). Furthermore, in the special case when validating NWP output with DL observations based on scanning measurement strategies, the model fields should be filtered or averaged according to the scan strategy to ensure that the model turbulence is evaluated on the same spatial scales that the DL actually observes.

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. Especially for TKE and EDR the retrieval method is well validated against in-situ tower data (Wildmann et al., 2020). Given that these turbulent properties are 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), this dataset is particularly suitable . 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 evaluationthanks to its consistency and comprehensiveness. , 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.

In this paper, we leverage the 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) to evaluate two turbulence configurations of the ICON model - the operational 1D TKE scheme, and a 3D Smagorinsky (Smagorinsky, 1963) scheme - for horizontal mesh sizes ranging from 2.1 km to 78 mfor 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 Section 2 we describe the DL measurement and retrieval method and the two model configurations. The currently operational ICON 2.1 km limited area configuration is evaluated 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 Section 3as a starting point. In Section ?? we focus on comparing the two model configurations and performance across horizontal resolutions. . A more in-depth validation of the parameterized turbulent length scale can be found in Section ??4. A case study with focus on model deficits w.r.t. excessive mixing is discussed in Section 5. The relevance of our results for subsequent NWP applications is explored with two examples in Section 6, and we conclude our discussion in Section 7.

2 Data and methods

2.1 Doppler lidar wind and turbulence measurements

At the Meteorological Observatory Lindenberg – Richard-Aßmann-Observatory (MOL-RAO) a StreamLine DL from the manufacturer HALO Photonics Halo Photonics StreamLine (now HALO Photonics by Lumibird) is used Doppler lidar for simultaneous wind and turbulence profile measurements in the lowest 600 m of the ABL . The applied 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 50.3m and 90.3m 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 180ns pulse length and a variable focus, is positioned relatively close to the mast, with an average distance of approximately 80 meters. 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 (TKE = 0.5 ($\sigma_u^2 + \sigma_v^2 + \sigma_w^2$)) (Kropfli, 1986)($\mathrm{TKE} = 0.5(\sigma_\mathrm{u}^2 + \sigma_\mathrm{v}^2 + \sigma_\mathrm{w}^2)$, 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 seconds such that 25 full scans are completed within a 30 minute 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(NTCs).

When performing conical scans with such a comparably high azimuthal and temporal resolution, the number of laser pulses per radial wind measurement is necessarily reduced. limited to $N_a = 2000$ due to the relation $N_a = \Delta\theta f_p/\omega_S$ (Banakh and Smalikho, 2013) . Here $\omega_S = 5^\circ\,\mathrm{s}^{-1}$ and $f_p = 10$ 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.

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

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 configurationsconfiguration

We investigate limited area simulations at five resolutions: The starting point is 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. The only difference to the operational forecast is the ICONmodel version (icon-2024.01-1) chosen to be consistent across all experiments described in this paper. The forcing setup for the sub-kilometer scale simulations (hereafter SKM) follow the methods described in Heinze et al. (2017). The simulations at the coarsest resolution of our setup (626 m mesh size, DOM01) are forced by the ICON operational forecasts at 2.1 km mesh size , while three two-way nested domains (DOM02Given the triangular grid used by ICON, DOM03, DOM04) are used with sequentially decreasing mesh sizes of 313 m, 156 m and 78 m. The inner-most domain size is about 22 km in the N-S direction and 24 km in the E-W direction. While the D2 setup uses 65 vertical levels, we use 120 levels for the SKM simulations. 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 parameterizations of convection, SGS orographic effects and gravity wave drag are switched off. For the land-surface component, we use the TERRA land surface model (Schulz and Vogel, 2020) with the option for subgrid land-surface variability turned off due to the high SKM horizontal resolution. The SKM setup includes the static land-surface data as used in the operational ICON D2 configuration , with addition of the ASTER orography data with the original resolution of 1 arc second (ASTER, 2019) and Corine land cover dataset at 100 m resolution. This setup is in fact identical to the one described in Sakradzija et al. (2025). The model binary is based on the ICON release tag configuration is comparable to the operational setup, except a more recent model binary ("icon-2024.01-1icon-2026.04") is used, which is available from the ICON Open Source git repository (https://gitlab.dkrz.de/icon/icon-model/). Simulations are initialized at 00UTC from the operational analysis and integrated over 36h. 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.

The two model configurations used in this study differ mainly in the schemes used to parameterize the turbulent transport and surface exchange: The NWP configuration (hereafter 1DCONF) employs the standard ICON turbulence scheme Turbulence is parameterized in ICON with the "Turbdiff" (Raschendorfer, 2001; Raschendorfer et al., 2003; Doms 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)and a subgrid-scale cloud cover parameterization. The second, LES-like configuration (hereafter 3DCONF) uses the 3D turbulence scheme of Smagorinsky-Lilly (Smagorinsky, 1963; ?; Dipankar et al., 2015) without cloud scheme. Importantly, the surface exchange in 3DCONF differs from that in 1DCONF and . The scheme is based on Louis (1979). Turbdiff started as a classical 1D TKE scheme with 2.5 order closure . The scheme has been extended over the years to include local horizontal shear terms contributing to TKE production, and scale transfer terms to account for extra turbulenceproduction caused by 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 $q := \sqrt{2\,TKE}$. The TKE equation, reproduced here in updated form from Goecke and Machulskaya (2021, their equation 9), takes the form:

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

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 SGS flows, such as horizontal shear from organized flows or local circulations induced by SGS orography. Additionally, the turbulent length scale is considered as mesh-size-dependent for higher horizontal resolutions. These feature are aimed at making Turbdiff more suitable for application in the turbulent gray zone. The ability to incorporate scale-transfer terms into the Turbdiff's TKE equationis due to the introduction of a generalized scale-separation concept, which divides the SGS turbulence spectrum into two parts. Here, the turbulence contributions of unresolved non-turbulent motions start at the maximum dimension of a grid cell D_g and extend down to a defined separation scaleL_p^{max}, which marks the transition between turbulence generated by organized motions and the "true" isotropic turbulence of the inertial subrange. If this upper part of the SGS turbulence spectrum ranging between D_g and L_p^{max} were not taken into account, a significant source of turbulence would be missed. As a result, the SGS turbulence would not be fully represented, and key 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 sections 1 and 2 of Goecke and Machulskaya (2021). Thus horizontal shear is accounted for through two different pathways, by including horizontal shear of the transfer from organized flows into isotropic turbulence would be underestimated.The positive effects of considering such transfer terms have already been demonstrated by a significant improvement in the NWP skills of EDR forecasts for aviation (Raschendorfer and Barleben, 2014). 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.

The parameterization of small-scale turbulence links the resolved and unresolved scales of motion. Its feedback on the resolved flow appears as additional source terms in the prognostic first-order model equations representing the turbulent flux divergence. In the current version of ICON only the vertical turbulent flux divergence is included in the prognostic wind equations, while the horizontal flux divergence is neglected. Therefore, neither the 3D turbulence contributions in 3DCONF nor the additional 3D-like extensions in 1DCONF are being exploited to their full potential. Turbdiff uses the Blackadar (1962) vertical length scale

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

with $l_m = \kappa \min\{\mathrm{D_g}, \mathrm{L_p^{max}}\}$, z_0 the roughness length and D_g the horizontal mesh size. κ denotes the von-Karman constant and $\mathrm{L_p^{max}} = 500\,\mathrm{m}$. $\mathrm{L_p} = \min\{\mathrm{D_g}, \mathrm{L_p^{max}}\}$ 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 $\mathrm{L_p^{max}}$ 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_\mathrm{h}$$K_m$ and K_h, play a key role in coupling the SGS turbulence to the resolved-scale flow. In the 1DCONF and 3DCONF, the These exchange coefficients are formulated differently. In 1DCONF they are determined prognostically by solving the TKE equation, while in 3DCONF they are considered diagnostically and are derived directly from the resolved shear. A brief overview of the key turbulence parameters used in both parameterization schemes is provided below and summarized in Table ??. defined as $K_m = q\ l_B\ f_m$ and $K_h = q\ l_B\ f_h$. The stability functions f_m and f_h are those described in Buzzi et al. (2011). A minimum value for the exchange coefficients of $K_{m,min} = 0.75$ and $K_{h,min} = 0.4$ is applied to maintain some degree of mixing under stable conditions.

The two schemes make slightly different assumptions about the turbulent length scale l used to describe the size of the most energetic SGS eddies. Turbdiff uses the classic Blackadar (1962) length scale

$$l_\mathrm{B} = \frac{\kappa z}{(1 + \frac{\kappa z}{l_m})}$$

with $l_m = \kappa \min\{\mathrm{D_g}, \mathrm{L_p^{max}}\}$ and by setting $D_g = 0.5(\mathrm{dx\,dy})^{1/2}$ where (dx dy) denotes the area of the triangular grid cell. κ denotes the Karman constant and $L_p^{max} = 500$ m. The Smagorinsky scheme instead uses

$$\frac{1}{l_\mathrm{S}^2} = \frac{1}{(C_\mathrm{S}\Delta)^2} + \frac{1}{(\kappa z)^2}$$

blending the same linear dependence on height (κz) as used in Turbdiff with a length scale based on the grid cell volume $(C_\mathrm{S}\Delta) = C_\mathrm{S}(\mathrm{dx\,dy\,dz})^{1/3}$, where (dz) denotes the layer thickness and $C_\mathrm{S} = 0.23$ the Smagorinsky constant. Since the layer thickness (dz) is significantly smaller than the horizontal mesh size for the coarser domains, 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 asymptotic length scale used in the Smagorinsky scheme is smaller than the Blackadar scheme, with the difference increasing for larger mesh sizes. For completeness, it should be mentioned that the height above ground z in Turbdiff also takes into account the roughness length and is slightly modified over forest canopy. However, for the Lindenberg site and 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 following discussion, this small deviation from the plain height above ground is irrelevant. In short, both length scales are resolution-dependent, but neither is explicitly stability-dependent. 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.

Turbdiff, being a TKE turbulence scheme, predicts the SGS-TKE as a prognostic variable. The Smagorinsky scheme on the other hand does not provide a direct estimate of the SGS-TKE. Underlying the Smagorinsky scheme is Given the Doppler lidar's elevation angle of $35.3°$, the assumption that turbulence generation via shear production and buoyancy production/destruction is balanced by dissipation. Thus, we can estimate the eddy dissipation rate $\varepsilon = -K_\mathrm{h}N^2 + K_\mathrm{m}S^2$, where the shear term S^2, Brunt-Väisälä frequency N^2 and turbulent diffusivity coefficients K_m 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 Section 2.1.

Given the horizontal mesh size of the ICON D2 configuration (approx. 2.1km), 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 30min time window. This calculation is performed during the model run using winds at every 20s time step, applying the method proposed by Welford (1962) for calculating incremental variances. In the following explanation, this 30min temporal variance will be denoted by σ_t^2, and K_h for momentum and heat are calculated by the Smagorinsky scheme. Note that in stable conditions with a Richardson number exceeding the Prandtl number (set to $Pr = 0.333$) the coefficients default to $K_{\mathrm{m}(min)} = 0.001$ and $K_\mathrm{h} = 0$ in this implementation. The SGS-TKE contribution can then be estimated using the relationship $\varepsilon = c_\varepsilon \frac{TKE_\mathrm{SGS}^{3/2}}{l}$ with $l = l_\mathrm{S}$, following de Roode et al. (2017). In Turbdiff a similar formulation is used to describe the EDR source term in the TKE equation by using $l = l_\mathrm{B}$. Note that L_V in Eq. (1) describes the size of the most-energetic turbulent eddies of the complete atmospheric turbulence spectrum while l_B and l_S describe the most energetic eddies only of the SGS part of the turbulence spectrum that is parameterized in the model (and therefore depends on the mesh size). 30min average wind at each grid point by u_t, where each of these variables can stand for any of the three wind components.

In Turbdiff, c_ε is chosen as $c_\varepsilon = 0.1704$. While this parameter should be a universal constant, it is in fact treated as a tuneable parameter within Turbdiff (Doms et al., 2021). Since no explicit assumption is made in the Smagorinsky implementation, we assume the same definition for the 3DCONF derivation of the 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 30min average winds u_t across all M grid points falling within the scan diameter:

$$\sigma_\mathsf{s}^2 = \frac{1}{\mathsf{M}-1}\sum_{\mathsf{i}=1}^{\mathsf{M}}(\mathsf{u}_{\mathsf{t,i}} - \overline{\mathsf{u}_\mathsf{t}})^2 \tag{4}$$

Here,

$$\overline{\mathsf{u}_\mathsf{t}} = \frac{1}{\mathsf{M}}\sum_{\mathsf{i}=1}^{\mathsf{M}} \mathsf{u}_{\mathsf{t,i}} \tag{5}$$

denotes the spatial average of the 30min 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:

$$\sigma_\mathsf{total}^2 = \overline{\sigma_\mathsf{t}^2} + \sigma_\mathsf{s}^2 \tag{6}$$

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

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

The SGS-TKE for consistency. representative of the 30min 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 30min 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.

Summary of the key turbulence parameters in 1DCONF and 3DCONF configurations as described in Sect. 2.2. Prognostic variables predicted by the scheme are bold. $f_\mathrm{s}^\mathrm{1DCONF}$ and $f_\mathrm{s}^\mathrm{3DCONF}$ denote stability functions and $q = (2\ TKE_\mathrm{SGS})^{1/2}$. 1DCONF 3DCONF subgrid-scale TKE $TKE_\mathrm{SGS} = (\frac{\varepsilon l_\mathrm{S}}{c_\varepsilon})^{2/3}$ eddy dissipation rate $\varepsilon =$

$c_\varepsilon \frac{\mathbf{TKE}_{\mathrm{SGS}}{}^{3/2}}{l_\mathrm{B}}$ $\varepsilon = -\mathbf{K}_\mathrm{h}\mathbf{N^2} + \mathbf{K}_\mathrm{m}\mathbf{S^2}$ integral length scale $l_\mathrm{B} = \frac{\kappa z}{(1+\frac{\kappa z}{l_m})}$ $\frac{1}{l_\mathrm{S}^2} = \frac{1}{(C_\mathrm{S}\Delta)^2} + \frac{1}{(\kappa z)^2}$ eddy diffusivity for momentum $\mathbf{K}_\mathrm{m} = q\, l_\mathrm{B}\, f_\mathrm{s}^{\mathrm{1DCONF}}$ $\mathbf{K}_\mathrm{m} = 2\, l_\mathrm{S}^2\, \rho\, \mid S \mid$ $f_\mathrm{s}^{\mathrm{3DCONF}}$

In both configurations, the parameterizations provide an estimate only of the SGS contribution of the TKE, but not the GS one. The reader is referred to Appendix A for a detailed description of how to calculate the latter, including a discussion of the The relative importance of contributions from temporal and spatial variability at different resolutions. Unless otherwise stated in the remainder of this paper, TKE refers to the total TKE, which is the sum of the GS-TKE and SGS-TKE. the GS and SGS TKE contributions will be explored in more detail in Sec. 4.

3 Evaluation of ICON operational limited area configuration

Comparison of time-height curtains for (top row) wind speed, (middle row) turbulence kinetic energy (TKE) and (bottom row) eddy dissipation rate (EDR) from (left) DL retrieval (a-c) and from (right) ICON simulations with D2 configuration (d-f). Each DL-derived dataset was subject to a specific quality flag (QC), resulting in differences in data availability.

A 5-day (8-12 June 2023) comparison of DL-based wind speed (a) and turbulence variables (b,c) at a height of 95.3 m with ICON simulations (D2 configuration) at a height of 77.0 m (the closest model height). The histograms also show the relative errors between the two, where vertical dashed lines indicate the $\pm$ 30 % range. Displayed USA-1 sonic (Metek GmbH) data from a 99 m tall mast 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 caused by the quality check procedure, which takes into account erroneous sonic data due to wake effects at the mast.

For the validation of selected components 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-hour simulations initialized at 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. Model results from the closest grid point to the DL measurement site are evaluated. DL based wind speedProfiles for wind speed, wind direction and turbulence (TKE, EDR) profile measurements retrieved from DL for this period are shown in Fig. 1[a]-[cd]. The period chosen encompasses five days with a typical summertime convective boundary layer evolution during daytime, while low level jets (LLJ) were observed during nighttime. Winds On several days, intense convective cells and precipitation developed, some of which produced cold pools. Wind speeds (Fig. 1 [a]) 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. 1 [c] 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 600m 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 June the 8th and the 10th. These are recognizable by a sudden change in wind speed, Depending on the case, the signature of the cold pools in the wind field is dominated either by strong directional shifts as on June 8th or by pronounced increases in the wind speed as on June 10th. With regard to the turbulence variables, cold pools are recognizable by increased TKE and a rapid collapse of the mixed layer, which varies depending on the individual cold pool. 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, FESSTVaL) 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 D2 model configuration model (Fig. 1 [de]-[fh]). Noticeable deviations occur in situations with Deviations are noticeable in situations where cold pool events are either observed or predicted. Inspection of 2 m temperature maps (not shown) confirm the occurrence of cold pools weaker, more distant cold pool in the model not only on June 8th and a stronger cold pool initiated in close proximity to the Lindenberg location on the 10th, but also on the 9th. Exact matches in terms of the timing, intensity and location of cold pools should not be expected . At coarser resolutions, the sharp gradients at cold pool fronts cannot be fully resolved. But even at higher resolutions, the since the development of deep convective cells and associated downdrafts is not deterministic, such that timing and location of events can vary. Interestingly, on the The night of June 9th the occurrence of the simulated but in reality unobserved cold pool is followed by an enhanced LLJ during the night which extends 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 levels and is also associated with a stronger turbulent mixing in ICON, which does not occur to this extent in the observations. Further studies on this topic A more detailed discussion of this case can be found in SectSec. 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 Fig. 2 and Fig. 3, respectively. Here, the time series plots not only provide comparisons of DL based wind and turbulence data at approximately 90 95 m height with corresponding ICON simulations at approximately 77m (the closest model height, level height) but also include independent sonic measurements that serve as a reference for estimating the DL retrieval quality. The direct at 90m 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. 2 [a]). Additionally, these time series comparisons clearly show the quantitative effects of the above described erroneous LLJ simulations on June 9th, i.e. the strong overestimation of wind speeds close to the ground. For TKE (Fig. 2 [b]), 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, usually by -30 %. At night, on the other hand, the while at night, the model overestimates the TKEwith values of up to 150 %. 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 TKE due to pulse averaging is effective (see SectSec. 2.1). The same cannot be said for the DL-based EDR retrieval In contrast, with maximum values between -100% and +400% the range of relative errors for EDR is considerably larger than for TKE (Fig. 2 [c]) . To date, a correction to compensate the underestimation of the DL-based EDR is unavailable for the EDR retrieval yielding to a systematic EDR underestimation relative to the sonic retrieval, which is consistent with the findings in Wildmann et al. (2020). Therefore, the comparison between DL-based EDR and ICON-based EDR gives the false impression of a fundamental overestimation of the model EDR over the entire observation period. In contrast, the comparison between sonic EDR and model EDR clearly shows that an overestimation of the model EDR occurs primarily at night, but daytime values appear to be reasonablewhich, 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 in the D2 configuration and DL 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 daytime (the day and at the transition from day to evening (ET) (Fig. 3 [a2],[a3]). At night between 21:00 and 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. 3 [a4]). 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 is still evident for disappears during the morning transition (MT) period, between 03:00 and 09:00 UTC (Fig. 3 [a1]), but . At the same time, the LLJ features in the model profile are becoming less well-definedpronounced, pointing to an earlier LLJ decay. The vertical wind gradients in the lowest 300 m are in better agreement , however. During the evening transition (ET) between 16:00 and 20:00 UTC, Similar to the wind profiles, in the TKE profiles better agreement can be observed during the day and during the ET (Fig. 3 [a3b2]shows , [b3]). By contrast, a systematic overestimation of the ICON wind profiles and a narrower IQR which indicates a systematically premature onset of LLJ formation with simultaneously lower profile variability. The TKE profiles show a slight, but systematic model underestimation during daytime and ET, but an overestimation at night TKE is observed at night, which changes to a systematic underestimation at the transition from night to day (Fig. 3 [b4], [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 daytime model overestimate of the EDR (Fig. 3 [c2]) 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. 3 [c3]). Together with the slight TKE underestimate for the same period, this is consistent with an early stabilization and LLJ-onset During the MT and daytime the model tends to overestimate the EDR (Fig. 3 [c1], [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 turbulent 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 to rapidly dissipated in the model. Interesting to note is the , 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. 3 [c4]). Particularly the stronger observed vertical decrease of EDR at night and in the early morning hours not reproduced by the model is striking, while the slopes of the TKE profiles are more comparable for the same time periods. 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.

To summarize, the D2 configuration of the ICON model not only qualitatively, but also quantitatively reproduces the turbulent properties of the daytime lower ABL well. At night, 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 LLJ 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, indicationg indicating a continuity of challenges across model generations (Finn et al., 2020). With this baseline model assessment based on the operational D2 configuration in mind, we will now explore to what degree increased resolution impacts the model performance.

4 Impact of increased model resolution for two different turbulence configurations

As the model resolution increases to the SKM range, the simulated wind and turbulence profiles are generally expected to capture finer-scale variability, thereby more closely approximating the observed profiles. SKM resolutions place numerical models in the gray zone of turbulence representation, where some of the variability is explicitly resolved, while other parts are still parameterized. The handling of this partitioning between resolved and subgrid turbulence can be different depending on the physical conception of the turbulence scheme (Honnert et al., 2020; Hope et al., 2024) which can also affect the overall model response - including quantities such as the total TKE, EDR and wind speed. In Sect. **??**, we use the singular value decomposition (SVD) as a practical tool for efficiently and objectively comparing the individual simulations to obtain a first compact overview of the differences between all simulations. In Sect. **??**, we then examine selected profile plots The relationship between LLJ winds, TKE and EDR will be explored in more detail which offers more targeted information on specific, localized differences. in the following section and in the case studies discussed in Sec. 5.

Spectra obtained from a singular value decomposition (SVD) of 2D time-height cross sections representing wind and turbulence profiles shown in Figs. **??** and **??** - **??** for the four different model resolutions DOM01 - DOM04 (x1- x4) and the two model configurations 1DCONF and 3DCONF. Corresponding SVD analysis results of DL data based on Fig. 1 are also shown.

3.1 Singular value spectra analysis

4 Resolved and unresolved scales

The SVD method has been introduced by Dvorský et al. (2010) as a suitable technique to analyze similarities and differences in two-dimensional (2D) data visualizations. The main advantage of SVD is that it can represent a 2D dataset with only a few dominant modes, capturing the dataset's essential structures or patterns. These modes are calculated by considering a 2D discrete data set as a matrix, **M**, and subsequently employing matrix decomposition

$$\mathbf{M} = \mathbf{U}\sum\mathbf{V}^T \quad ,$$

where $\mathbf{U}$ and $\mathbf{V}^T$ contain patterns describing the variation along the rows and columns of $\mathbf{M}$, respectively, while $\sum = diag(\sigma_1, ..., \sigma_r)$ is a diagonal matrix showing the importance of each pattern. The sorted entries of the latter provide a singular-value-spectrum where the singular value of the first mode, i.e. σ_1 , captures the dominant large-scale structure of the 2D field. Successive modes σ_i represent increasingly smaller-scale or more localized patterns. Since the diagonal matrix contains all the essential features of the original 2D data set $\mathbf{M}$ it thus serves as a kind of "fingerprint" of this field. Hence, SVD reduces the dataset's complexity to a small number of key structures, facilitating comparison between large numbers of 2D datasets. Note that, since SVD spectra reflect the relative variance captured by each mode within a given dataset, comparing such spectra can only detect differences in variability patterns, not absolute differences.

The analyses shown in Fig. **??** use the SVD method to compactly compare time-height cross sections of simulated wind and turbulence variables at the observation site, alongside the relevant DL observation data, in order to quantify the similarity or difference of the underlying structures. This is done for each model configuration (i. e. 1DCONF and 3DCONF) and across all resolutions (i.e. DOM01 - DOM04). To enable a fair comparison between model and observation, the DL data were vertically subsampled to match the vertical resolution of the simulations for this analysis.

A first glance at the spectra comparisons in all panels of Fig. **??** reveals some significant differences between the two model configurations. This is evidenced by diverging spectra at all resolutions, although the degree of divergence varies depending on the mesh size and the atmospheric variable. For the time-height cross section of wind speed (Fig. **??**, top row), structural similarities can be observed for the first three modes reflecting larger scales or structural patterns. Differences are limited to the range of smaller local scales, which is indicated by the diverging curves from mode four onwards. It is striking that the relative position and energy level of the 3DCONF and 1DCONF related spectra hardly change depending on the model resolution. In other words: There are systematic differences in the windspeed patterns at smaller scales between the two model configurations, but increasing resolution does not much alter the model wind speed.

For the The evaluation in the previous section focussed on wind, total TKE and EDRcross sections, the spectra show greater sensitivity to resolution, particularly for the 3DCONF, and substantial differences between model configurations exist for all spectral modes, particularly at coarse grid spacing. The agreement between the 3DCONF spectra and the spectra observed by DL is poor at coarser mesh sizes, but improves at smaller ones. This behaviour should come as no surprise and highlights that the underlying turbulence scheme is primarily suitable for LES applications and is not well adapted to coarser gray-zone resolutions. For the 1DCONF, the TKE and EDR spectra remain quite consistent across all resolutions, and in good agreement with observations up to mode 6. This suggests that efforts to improve the representation of SGS turbulence by introducing a generalized scale-separation concept (Sect. 2.2) have been quite successful.

The above results clearly reveal that both the model resolution and the choice of the turbulence parameterization have a significantly stronger impact on the turbulence variables than on the wind velocity itself. This gives evidence that the simulated wind fields are more strongly governed by large-scale dynamics such as pressure gradients making them on these scales less sensitive to changes in the turbulence schemes and model resolution. The lack of improvement in the small-scale local wind field structure with increasing model resolution, however, allows different conclusions. Either this is due to the model resolution, which even for DOM04 might still be too coarse for a realistic simulation of smaller local structures, or the turbulence effects might be ineffective due to model deficiencies in the coupling between parameterized turbulence and directly simulated wind fields. This could be a plausible explanation, since neither in the presence of structural changes in the turbulence fields, as in the case of the 3DCONF simulations, nor in the absence of structural changes, as in the case of the 1DCONF simulations, there are structural changes in the wind field in response to the turbulence schemes. This explanation can be further supported by the fact that in the used model formulations for both configurations horizontal divergence terms of the turbulent fluxes are not yet included in the first-order momentum equations (Sect. 2.2) . These terms, however, might become increasingly important in the gray zone where horizontal differences are better resolved, alongside vertical flux divergence terms. As a consequence of neglecting these terms, the additional horizontal flux information provided by the 3D turbulence parametrization schemes does not feed back into the horizontal momentum tendencies, and the simulated wind fields remain essentially unchanged, even at higher resolution. Similar conclusions have been drawn in Doubrawa and Muñoz-Esparza (2020) by investigating the gray zone performance of the Weather Research and Forecasting (WRF) model.

Interestingly, Goger and Dipankar (2024) performed a similar comparison of the two ICON configurations for the Inn valley in the Alps. In their study, an increase in horizontal resolution led to an improved representation of winds in the valley, and the 3DCONF shows advantages over the 1DCONF for the timing of the up-valley flow and its vertical structure. This is attributed to the relatively more important role of horizontal shear around orography starting at relatively coarse resolutions of 1km that may be better captured by the three-dimensionality of the 3DCONF. In contrast, the surface heterogeneity (and related potential for generating horizontal shear) in our SKM configurations does not change significantly with smaller mesh size as surface properties are interpolated from the coarser ICON forcing at 2.1 km mesh size. Prescribing surface heterogeneity on the scale of the resolved mesh size may be a missing prerequisite for realizing the full benefit of the finer mesh over flat topography.

Finally, another important point to note is the fundamentally higher energy level of the 3DCONF spectra compared to the 1DCONF spectra for all variables. Theoretically, this means a higher time-height variability in the wind and turbulence variables, e.g. due to stronger gradients, stronger gusts and more pronounced turbulence structures. However, since the DL-based spectra show better agreement with the 1DCONF-related spectra, this higher variability in the 3DCONF-based simulations does not appear to be realistic and must be critically scrutinized. The latter supports the use of the 1DCONF in the new ICON 500 m configuration.

4.1 Time-height cross sections and profiles

Time-height plots for wind and turbulence variables in analogy to Fig. 1 but with focus on a comparison of ICON SKM simulations with the two configurations 1DCONF (a- c) and 3DCONF (d- f) on the finest mesh size DOM04.

Same as in Fig. 2 but for simulations with the ICON SKM setup on domain DOM04. Results from both turbulence configurations 1DCONF and 3DCONF are shown.

Same as in Fig. 3 but for simulations with the ICON SKM configuration on domain DOM04. Results from both turbulence configurations 1DCONF and 3DCONF are shown.

Figure **??** shows time-height cross sections of winds and turbulent properties analogous to Fig. 1 for both SKM model configurations using the highest model grid resolution, i.e. 78 m mesh size (DOM04). For brevity, the remaining set of figures for DOM01-DOM03 have been moved to Appendix **??** (Figs. **??** - **??**). A visual comparison of these figures corroborates the previous findings from the spectra, revealing distinct differences between the two turbulence configurations (1DCONF . This section analyses the contributions from the individual SGS and 3DCONF), but only minor changes are observed within each configuration across the different spatial resolutions. For this reason, the following direct profile analyses concentrate on comparing DOM04 simulations. To better assess potential benefits of the higher spatial resolution (and the choice of the turbulence scheme) compared to the D2 configuration we additionally use the graphical representations shown in Figs. **??** - **??**, which are consistent with those discussed in Sect. 3. However, rather than repeating the full comprehensive evaluation against observations as done in Sect. 3, the focus here is restricted to key aspects where significant deviations or improvements compared to the coarse-resolution simulations are observed.

Consistent with the singular-value-spectra (Fig. **??** a4), we do not see significant differences between 1DCONF and 3DCONF in the main patterns of the time-height evolution of the wind speed (Fig. **??** a, d). The small-scale local differences indicated by the singular value spectra become visible, for example, in the slightly different vertical extent of the LLJ towards the earth surface during the night of June 9th GS components (as discussed in Sec. 2.3) to 10th. At $\sim$90 m height this contributes to a somewhat better agreement with DL-based wind speed observations for the 3DCONF (Fig. **??** a). Despite the largejump in horizontal mesh size from the total TKE across the diurnal cycle. With a mesh size of 2.1 km to 78 m, no systematic improvements of the simulated winds are evident for either configuration when comparing Figs. 1 a, d, **??** akm, d, and Figs. 2 a,**??** aand Figs. 3 a1-a4, **??** a1-a4. The error statistics show that the frequency of strong over- or underestimates in wind speed of up to $\pm$ 100 % remain similar even with 78 m mesh size.

The time-height cross sections of the turbulence variables TKE and EDR show more noticeable differences, particularly at night and in the early morning hours (Figs. **??** b,cand e,f). 1DCONF produces the highest TKE values near the surface, rapidly decreasing with height, while the 3DCONF provides an elevated relative TKE maximum in heights between 100 m and 200 m with lower values near the surface. Although the DL observations (Figs. 1 b, c) do not quite reach the surface, the 1DCONF TKE confined to a shallower night-time turbulent layer matches the observations more closely. This is also evident in the composite TKE profile shape presented in Figs.**??** b1, b4. While the nighttime composite TKE profiles for the coarser D2 configuration (Fig. 3 b4) show a slight overestimate, the higher resolution DOM04 composite profiles now show a slight underestimate at night and in the early morning hours for the 1DCONF. While our previous consideration of TKE based on SVD showed generally stable total TKE across scales for the 1DCONF, this new perspective based on composite profiles suggests a slightly increasing underestimation of total TKE as the mesh size becomes finer. Supporting evidence for this is provided in Sect. **??**. In absolute terms, the 3DCONF with slighly higher TKE values gives a better match to observations here, despite a more unrealistic vertical distribution. However, given that the SVD spectra for TKE revealed larger singular values than the observations, indicating either incorrect distribution of turbulence kinetic energy or excessive production, it is concluded that the apparent agreement in total TKE is likely coincidental rather than physically consistent .

There is no impact of the model resolution on the positive model bias in the EDR profiles at night and in the early morning hours (Fig. 3 c4and Fig. **??** c4). This means that in the 1DCONF and with increasing model resolution there is a tendency at night that less turbulence is generated but more energy dissipated compared to the observations. Taking into account that τ = TKE/EDR represents a characteristic time scale defining a measure how persistent or intense turbulence is, this indicates shorter-lived turbulence structures in ICON compared to the real atmosphere. Since dissipation remains a SGS process irrespective of the model mesh size in our simulations, it does not surprise that the model's EDR bias remains unchanged across resolutions.

Of particular note are the nights of June 09th/10th and 10th/11th, when both model configurations have stronger vertical mixing than observed (i.e. enhanced TKE and EDR) up to altitudes of 600 m (Figs. **??** b, cand e, f), similar as observed in the ICON D2 simulations (Figs. 1 econfiguration is on the threshold of the turbulent 'gray zone', where large, f). Since this is not only in contrast to the DL based measurements (Figs. 1 b,c) but also in contrast to the other nights during the considered 5-day time period a more detailed case study analysis on this modeling deficits is provided in Sect. 5. 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.

505 **5** **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 Appendix A) to the total TKE to evaluate the model's ability to adjust SGS turbulence appropriately. This 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 integral 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 many turbulence parameterizations rely 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 existing parameterizationsthe Turbdiff parameterization scheme, especially under conditions with stable stratification and strong shear.

4.1 Scale partitioning of turbulent kinetic energy

In an ideal scale-adaptive turbulence closure model, the SGS-TKE should decrease as the resolution increases, while the GS-TKE should increase. This ensures that the total TKE remains approximately constant, thereby maintaining energetic consistency. Figure 4 illustrates how the partitioning The time-height plots in Figs. 4 [a], [b] show the breakdown of the total TKE at a height of 87.5 m (Fig. **??** as shown in Fig. 1 [bd]) between the SGS- and 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 depends on the model resolution and time of day for both 1DCONF and 3DCONF. The expected decrease in SGS-TKE and increase in GS-TKE across grid resolutions from DOM01 to DOM04 can be seen for all times of the day for both configurations. However, the increase in GS-TKE is not always consistently strong enough to compensate for the decreasing SGS contribution. This results in a tendency of a systematic decrease in total TKE with increasing resolution, particularly during the night and morning transition period which is in agreement with the findings discussed in Sect. **??**. During the day and the evening transition the total TKE tends to be well maintained across resolutions, thanks to the GS-TKE's strong compensating power. This different behaviour might be explained by the current coupling between the turbulence closure and the first-order momentum equation (Sect. 2.2). In the momentum equation the divergence terms of the fluxes are key to the redistribution of SGS- and GS-TKE. They act as the bridge that transfers momentum and energy from the parameterized SGS motions into the explicitly resolved wind field. Thus, the variability in the wind fields increases due to two factors: the finer mesh sizes, which explicitly resolve smaller-scale motions, and the energy transfer from the SGS-TKE through the flux divergences. If parts of this transfer pathway are missing, such as the horizontal divergences, the decrease in SGS-TKE with resolution is not sufficiently compensated by an increase in resolved TKE and the total TKE consequently drops. Figure 4 might give an indication of this effect. Since vertical transport dominates in the convective boundary layer and the vertical divergence terms are retained in the momentum equation, the fluxes provided by the turbulence closure can effectively feed the GS-TKE through enhanced vertical variances. This ensures that the total TKE remains nearly constant across resolution in convective conditions (to total TKE was analyzed across the diurnal cycle in Fig. 4 [c2d]). In stable conditions, however, vertical turbulent transport is strongly suppressed by stratification, so horizontal transports become comparatively more important. Under these conditions, the divergence of horizontal momentum fluxes is one of the remaining mechanisms to redistribute SGS energy into the resolved flow. This feature, however, is not yet implemented in the first order momentum equation and thus might explain the slight loss in total TKE with finer mesh sizes (Fig. 4 . When viewed in conjunction with the time-height plot for the simulated wind in Fig. 1 [c4e]). it is apparent that the GS-TKE contribution is strongest during the passage of the simulated cold pool on June 10th or in association with the development of the nocturnal LLJs. The briefly occurring, high GS-TKE values extending from the surface to 500m 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.

Beyond the general issue of scale partitioning discussed above, the boxplots in Fig. 4 Finally, Figs. 1 [d1c1]-[d4c4] also provide an overview of the relative proportion of SGS-TKE to total TKE. They illustrate that , in addition to the expected dependence on resolution, the proportion of SGS-TKE also exhibits a pronounced diurnal cycle. A straightforward explanation is that during daytime convective conditions, a larger fraction of the total TKE is represented explicitly, since the dominant turbulent structures occur on spatial scales that are well resolved by the model grid. In contrast, under nighttime stable conditions, turbulence is confined to much smaller and weaker eddies that are not captured at the given mesh sizes, so a larger share of the present a statistical comparison of median SGS-, GS-, and total TKE remains in the SGS component. Only at the highest resolution, i.e. DOM04, the SGS fraction decrease slightly, suggesting that the grid spacing begins to capture the largest eddies that occur under stable nighttime conditions. The general differences between 1DCONF 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 and 350m 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 3DCONF indicate that the configurations partition turbulence differently into resolved and SGS components, reflecting their distinct scale-awareness and energy redistribution characteristicshere the SGS component alone already exceeds the observed total TKE. This suggests that the bias is more 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 SGS integral turbulent length scale l_X

In 1DCONF and 3DCONF, $l_X \in \{l_\mathrm{B}, l_\mathrm{S}\}$ only refer to the size of the most energetic SGS turbulent eddies (Sect. 2.2) and are, hence, not identical to the DL-derived L_V given through Eq. (1). The latter relates to the most energetic eddies of the full spectrum including both SGS and GS turbulencestructures (Sect. 2.1). As described in Section 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 definition for the models total TKE as estimation of the model's total TKE (i.e. the sum of the GS-TKE and SGS-TKE GS and SGS contributions) and the model definitions for EDR (Sect.2.2EDR 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 unnecessary not necessary here. Then, after some algebraic manipulation a functional relationship can be derived, reading

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

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 $l_X \in \{l_\mathrm{B}, l_\mathrm{S}\} l_B$. 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_X l_B also.

A comparison between DL measurements for 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. ??Figs. 1 [d], [f]), in Fig. 5 [a] all data are excluded where EDR $< 10^{-4}$ m^2 s^{-3}while in , while Fig. 5 [b] all observations are showndisplays all unflagged data. The time-height cross sections in Fig. 5 [a] 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 ($\sim$ 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 to 200 m. The large L_V values in the late afternoon reflect typical examples of NTCs discussed in Sec. 2.2.

Taking into account Considering the time-height cross sections shown in with unflagged observational data (Fig. 5 [b]), 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 June 8th (see SectSec. 3), whose passage involves non-stationary, rapidly evolving wind fields. This suggests that the formula is not applicable there in these instances because the underlying assumptions, e.g. isotropy, quasi stationary turbulence, existence of a well-developed inertial range (see SectSec. 2.1), are violated at these timesnot 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 sub-range. When Eq. (1) is employed within the framework of coarse-resolution numerical models, where the SGS turbulence exhibits strong anisotropy, all the uncertainties are usually attributed to the parameterization of the turbulence length scale which no longer scales with the inverse cut-off wave number. The length scale formulations for coarse-resolution models (e.g., NWP and climate models) should account for turbulence anisotropy and possibly also non-stationarity in a rational way. range.

Figures Figure 5 [c] - fshow 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. This is due to an inadequate parameterization of l_X as can be explained in more detail by means 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. 4 [d]. Beyond these episodes, the modeled length scales show little temporal evolution throughout the day. A closer examination of Eq. (8) . Especially if the and Fig. 5 [d] illustrates that the insufficient growth of the turbulent structures directly reflects the definition of l_B. If the SGS-TKE dominates the total TKE yielding $g < 1$, it can be shown mathematically that $L_V^* \sim c_4 c_\varepsilon^{-1} l_X L_V^* \sim c_4 c_\varepsilon^{-1} l_B$, implying that the way l_X l_B is parameterized dictates the behaviour of L_V^* through a direct proportionality relationship. Model conditions with $g < 1$ are naturally more prominent in DOM01 than in DOM04. This is shown in Figs. ?? This is graphically illustrated in the similarity between Fig. 5 panels [a- dwith each panel displaying the time-height cross section for l_X as parameterized in 1DCONF (see Eq. (3)) and 3DCONF (see Eq. (??)) with white areas marking where model conditions does not satisfy $g < 1$. Using DOM01 as an example, the time-height patterns of l_X shown in Figs. ?? a, c] clearly dominate the time-height patterns of L_V^* shown in Figs. 5 and [c, d]. This provides additional graphical evidence how l_X governs L_V^* if $g < 1$ and thus illustrates that the definition of l_X based only on fixed (grid) parameters does not accommodate a temporal evolution of L_V^* as observed for L_V , 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. 5 [ac] . Signals of larger turbulent structures during the day evolve only sporadically if model conditions with g >= 1 exist.

For DOM04, the GS contribution to the total TKE becomes more relevant (Figs. ?? b, d) such that L_V^* shows less of the imprint of l_X and can evolve (Figs. 5 d, f) in better for L_V^* > 500m during daytime. Yet we saw in Fig. 3 that the magnitude of the total TKE is generally in quite good agreement with the observations. The results of the simulations based on 1DCONF, however, appear more plausible than those based on 3DCONF. , 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 $L_p = l_B/\kappa$ marking the upper limit of the inertial range, and a stability-dependent formulation would better accommodate the temporal evolution of L_V^*values based on 3DCONF have an excessively large vertical gradient instead of a gradual temporal increase throughout the day , resulting in unrealistically large eddies > 1000 m within the lowest 300-600 m of the modeled ABL compared to the observations.

We can summarize that the diurnal variation of the length scale is better captured at higher resolution, primarily because the parameterized SGS turbulence becomes less important. A potential solution to improve the model performance at coarser resolutions would be to introduce a stability dependence of l_X. The validity of the EDR derivation based on the Kolmogorov relationship (Table **??**) would have to be re-evaluated based on whether the stability-dependent length scales restrict the turbulent spectrum to the fully isotropic range.

Time-height cross sections for the prescribed mixing length scales $c^* l_X$ with $l_X \in \{l_\mathrm{B}, l_\mathrm{S}\}$ as used in the ICON turbulence parameterization schemes (see Table **??**) for the time periods similar to the plots shown in Fig. 5. The white masked areas indicate regions where the GS-TKE contribution is larger than the SGS-TKE contribution to the total TKE.

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

All of the ICON simulations analyzed here, both in the D2 and SKM forcing configurations, exhibit the same striking behaviour: Building upon the preliminary finding in Sec. 3 that the strong overestimation of modeled near-surface wind speeds during the night of june 9th to 10th is closely related to excessive vertical mixingduring the nights of June 9-10 (night B) and June 10-11 (night C) throughout the lowest 600 m of the ABL. This becomes evident by comparing DL-derived EDR profiles shown in Fig. 1 , 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 Figs. 6 [ca]with both D2 simulations shown in Fig. 1 fand SKM simulations shown in Figs. **??** - **??** c, [fb] and Figs. **??** the EDR profiles in Figs. 6 [c], [fd]. Although model performance improves slightly from DOM01 to DOM04, unrealistic mixing remains significant at all horizontal resolutions and is more pronounced in 3DCONF than in 1DCONF. Possible causes that may be related to the parameterizations and its differences are examined and discussed in more detail in this section.

The general stronger vertical mixing in 3DCONF compared to 1DCONF can be attributed to the fact that the eddy diffusivity K_m in 3DCONF is directly proportional to the resolved shear $\mid S \mid$. In regions of strong LLJ shear, such as those observed during nights B and C, this can amplify turbulence production ($K_\mathrm{m} S^2$). The latter can then relatively easily overwhelm buoyant suppression ($-K_\mathrm{h} N^2$) even in strong stable stratification which may explain the higher EDR values calculated from the local equilibrium assumption between the production and dissipation of turbulence (see Table **??**). Note that the local equilibrium assumption provides a diagnostic approximation of the TKE equation and eliminates the need to solve for the prognostic evolution of turbulence. This is in contrast to 1DCONF where the TKE equation is solved explicitly, enabling K_m to evolve over time as a function of the predicted TKE. These differences in the treatment of K_m may explain the observed relative differences in enhanced mixing between 1DCONF and 3DCONF under stable, shear-dominated night-time boundary layer conditions. Another possible contribution may come from the previously mentioned differences in the surface transfer scheme employed in the two configurations. The role of the surface transfer vs. turbulent transfer is relatively greater under stable conditions. Unfortunately, the current configurations do not allow us to quantify how much the different surface transfer schemes contribute to the EDR differences between configurations.

However, the above arguments only help to explain relative differences in EDR overestimation between 1DCONF and 3DCONF. The fact that both configurations generally exhibit excessive mixing requires further explanation. This raises questions about the effectiveness of the stability-dependent damping functions f_s additionally included in K_m of each configuration, to take into account the influence of thermal stratification on turbulent mixing. Turbulence within the LLJ region is governed by the competing effects of shear and stratification. While shear production tends to generate turbulence, the strong stability in the nocturnal ABL suppresses vertical mixing. The stability functions in both configurations are desigend to represent this balance by damping K_m under stable conditions while remaining active under weakly stable or unstable conditions. In this context, the enhanced mixing during nights B and C could also be due to the insufficient effectiveness of f_s, preventing sufficient damping of shear induced turbulence and therewith EDR.

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 3DCONF, f_s is defined through the Richardson number Ri and the Prandtl number Pr_t, reading

$$f_\mathrm{s} = (1 - Ri/Pr_t)^{1/2} \quad \text{for} \quad 1 - \frac{Ri}{Pr_t} > 0$$

where Pr_t here defines a critical threshold for turbulence damping (Basu and Porté-Agel, 2006). Recall that in these experiments Pr_t is set to 0.333 implying $f_\mathrm{s}^2 < 0$ and $K_\mathrm{m} = 0$ in stable conditions if $Ri > 0.333$ (see Sect. 2.2) 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.

Time-height cross sections for f_s^2 as calculated in the SKM setup using 3DCONF for the 5-day analysis, are shown in Fig. **??**, alongside the associated cross sections for EDR A slightly different perspective on the same case study is presented in Fig. During daytime within the convective ABL f_s predominantly enhances turbulence. With increasing resolution from DOM01 to DOM04, f_s decreases slightly, which results in more realistic EDR values compared to the observations(Fig. 1[c]). This can be interpreted as another indication that the modified Smagorinsky - Lilly scheme is not well-suited for coarser sub-kilometer scales. Furthermore, it becomes apparent that f_s does not support a clear damping effect for all nights, despite pronounced LLJs occurring in the observations (Fig. 1 6 [ae]) suggesting strong stable conditions on all five nights. During nights B and C the occurrence of regions with $f_\mathrm{s} > 0$ is much more pronounced and has greater vertical extent compared to the other nights. These regions coincide with those of strong nighttime EDR overestimation, meaning f_s is less efficient at damping shear induced turbulence during nights B and C compared to the other nights. In contrast to the situation in the daytime ABL, f_s does not show significant changes from DOM01 to DOM04 at night. The previously mentioned deactivation of the turbulent flux horizontal divergence terms in the prognostic equations (Sect. **??**) may also be relevant in this context, as they apply not only to the winds but also temperature. The scale-invariance of f_s^2 (which indirectly depends on temperature) at night may be a reflection also of the weak coupling between turbulence scheme and thermodynamic resolved flow.

a-b: Time-height cross sections of the argument of the stability function (Eq. **??**) from SKM simulations with 3DCONF for mesh sizes based on DOM01 and DOM04 simulations. c-d: Same as in a-bbut for the eddy dissipation rate (EDR)

The stability function employed in the Turbdiff scheme to account for the impact of thermal stratification on K_m (Doms et al., 2021) is more complex than in the Smagorinsky-Lilly scheme, and an in-depth investigation would exceed the scope of the current work. However, since the error patterns in the EDR profiles between 1DCONF and 3DCONF are qualitatively similar, it is likely some of the conclusions drawn for the stability function in 3DCONF could also apply to 1DCONF and explain the similar EDR overestimation found in those simulations (Figs. **??** c,f).

As mentioned briefly above, according to Eq. **??**, the effectiveness of $f_s = f_s(Ri, Pr_t)$ depends on (i) the quality of the simulated thermal stratificationand wind shear, as these are directly involved in the calculation of Ri, and (ii) the choice of Pr_t. Consequently, any modelbias in these input parameters can propagate into erroneous stability functions, impacting the resulting turbulent fluxes. Errors in modeled turbulent fluxes directly affect the evolution of wind and temperature profiles in the ABL, which in turn modify Ri used in f_s. This creates a positive feedback loop: inaccurate Ri values further degrade the performance of f_s, leading to persistent biases in the vertical profiles of mean and turbulent variables. One example of this is the imprecise temporal evolution of the nocturnal LLJ in night B shown in Fig. **??** for the time period from 18:00 UTC to 21:30 UTC. For both model configurations the modeled profiles are subject to a relatively similar temporal evolution. The first four wind profiles between 18:00 UTC and 19:30 UTC describe an intensifying LLJ with an upward moving LLJ maximum (LLJ$_{max}$). Further intensification of the 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 $_{max}$ as it is visible in the DL wind profiles with the time stamps from 20:00 UTC to 21:30 UTC, however, is not observed. Instead, from this point onward, significantly higher eddy dissipation rates are modeled within the LLJ$_{max}$ region compared to DL based observations, which indirectly indicates excessive momentum fluxes. These momentum fluxes rather seem to contribute to transporting kinetic energy from higher altitudes towards the ground at the expense of further intensifying the LLJ$_{max}$, which then leads to increased wind amplification below 200 m, in contrast to observations. development. To resolve the observed near-surface wind biases and the lack of atmospheric decoupling, future adjustments to the ABL scheme's parametrization 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 Sec. 2.2).

Despite the above described positive feedback loop, an initial trigger is requiredto be activated. One obvious trigger could stem from inappropriate parameter settings including, for instance, Pr_t used in f_s. In 3DCONF this parameter is fixed throughout the entire day. This can be problematic, especially in the stable ABL, where the actual Pr_t is often significantly higher than 0.333. A stability-dependence of Pr_t is well known (Basu and Porté-Agel, 2006; Li, 2019) which underlines the necessity of its dynamic adaptation on atmospheric stratification. Furthermore, it is worth noting that previous work by Buzzi et al. (2011) on excessive mixing observed in the COSMO model , i.e. the conceptual and operational predecessor of ICON at DWD, already identified excessive mixing as a consequence of using relatively large minimum turbulent diffusion coefficients for K_m to artificially maintain turbulent diffusion. In this particular experiment using 3DCONF, this explanation can be ruled out because $K_{\mathrm{m}(min)} = 0.001$. The same conclusion cannot necessarily be drawn for 1DCONF, where $K_{\mathrm{m}(min)}$ is set to 0.75.

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-up follow-on applications can only be meaningfully assessed in the context of how the model results are intended to be used. The ICON simulations considered in this work show only modest improvements in wind speed, but somewhat more significant improvements in the turbulence variables with refined model grids. In this section we want to address the question whether these gains are sufficient to positively impact subsequent 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 Richardson number 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 DL based EDR time-height cross sections based on DL measurements in Fig.1 [c] show noticeably strong vertical gradients at certain altitudes. Under conditions of buoyancy driven turbulence these gradients indicate a separation of atmospheric layers with strong turbulence in the mixing layer from layers with very weak turbulence in the free atmosphere. 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 5-day 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 Richardson-number based Ri-based approaches.

a, d: Time-height cross sections of simulated EDR profiles using 1DCONF and 3DCONF with an additional white line indicating the MLH based on the threshold approach using $\mathrm{Thr_{EDR}} = 10^{-4}\ \mathrm{m^2\ s^{-3}}$ (see Appendix ??) . b, e: Time-height cross sections of simulated potential virtual temperature profiles using 1DCONF and 3DCONF with an additional black line indicating the MLH determined using the (offline) MLH diagnostic based on the Ri threshold approach with $\mathrm{Thr}_{\mathrm{Ri}} = 0.2$ c, f: Direct MLH comparison from the different threshold approaches against observations based on both radiosonde soundings and Doppler Lidar derived EDR profiles.

Figure 7 shows a comparison of results for diagnostically derived MLHs from ICON simulations based on 1DCONF and 3DCONF, respectively. Unlike in Figs. ?? Figs. 7 [ca] , and [fb] , the show time-height EDR cross sections in Figs. 7 across sections for the simulated virtual potential temperature (θ_v) and EDR, dextend up to 4000 m, showing the MLH evolution over the full day. While the Figs. ?? c, fclearly show systematically higher EDRvalues in the lowest region of the ABL for 3DCONF simulations, the Figs. respectively. The additional line drawn in Fig. 7 [a] , dshow the opposite at higher altitudes. This contributes to systematically lower EDR-based MLHs from 3DCONF simulations compared to 1DCONF simulations (see Figs. 7 cand f), especially in the daytime convective ABL. By contrast, the MLHs diagnosed purely based on vertical is a result of applying the threshold approach on profiles of the bulk Richardson number (Ri) reveal similar results for 1DCONF and 3DCONF (Fig. 7 b, e), reflecting comparable thermodynamic profiles (Fig. 7 b, e) in both model configurations. The diagnostic used here calculates the bulk Richardson number Ri_B) defined on each model level as

$$Ri_B = \frac{gz(\theta_v - \theta_s)}{\theta_s(u^2 + v^2)} \tag{9}$$

where θ_v, u and v are the virtual potential temperature and winds on the current model level, and θ_s is the surface virtual potential temperature (after Sørensen et al. (1996))(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 the same as used in the 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. When the results from two diagnostic methods are compared for each model configuration, it can be seen that during night , except those suffering from excessive mixing (see Sect. 5), the MLHs estimated from EDR profiles are in better agreement with those estimated from DL-based EDR profiles and from Radiosonde-based Ri profiles (see Appendix ??) than the MLHs using pure Ri-diagnostics, and are certainly more realistic than an assumed fixed value. In analogy with the original post-processing, during night the MLH is set to 276 m. Unlike in Fig. 1 [f], the time-height EDR cross section in Fig. 7 [b], extends up to 4000 m, showing the MLH evolution over the full day. The white contour in Fig. 7 [b] marks the diagnosed MLH using the EDR-threshold approach described in Appendix A.

Panel [c] in Fig. 7 then shows the two model diagnostics together with the MLH retrieved from RS and DL measurements. Especially during the night, the EDR-based MLH derived from ICON tends to overestimate the observed MLH. Two aspects contribute to this overestimate. As discussed in Sec. 3 and shown in Figures 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. However, this would not address the situation-dependent overmixing (and therefore overestimated EDR) showcased in Sec. 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). We propose the EDR-based approach shown above as a valuable alternative to complement Ri-based formulations under stable conditions. 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.

Within the convective ABL the assessment of the different MLH diagnostic approaches is more difficult. While for 1DCONF both MLH diagnostics show comparable results for the first two days of the 5day-period, the difference between them is quite large for the last two convective days. Note that it is exactly the other way around for 3DCONF. Furthermore, for The comparisons between EDR- and Ri-based MLHs over the last two days , a time shift in the temporal evolutionbetween EDR-based MLHs and Ri-based MLHs is evident for both configurations, indicating earlier growth and decay of 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. Finally it is worth pointing out that June 10th denotes a special case. The strong simulated cold pool event (see Sect. 3) contributes to a fast response in the Ri profiles towards stabilisation which abruptly lowers the MLH . The EDR profiles still keep turbulent mixing resulting in higher MLHs which is, at least for 1DCONF, in good agreement with the MLHs derived from radiosonde soundings

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

With the While mean wind speed is a primary driver in wind energy, turbulence intensity (hereafter TI) defined through

$$TI = 100 \ \frac{\sqrt{2 \ \ TKE}}{V_h} \quad \text{in} \quad [\%]$$

a combined quantity from turbulence kinetic energy and horizontal wind speed is obtained, which enables a normalized consideration of turbulence. This is of particular interest in the wind energy sector (Tai et al., 2023). 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 model represents this variable, and on its potential as an additional source of information for ambient TI.

Turbulence intensity predicted by the ICON model in the D2 and SKM setups is analyzed here to assess whether the simulations can provide reliable data for characterizing the ambient turbulence at a wind farm site. This background component is essential for site assessment, since it affects turbine loads and fatigue as well as the recovery of turbine wakes. In operating wind farms ambient TI must be complemented by the additional turbine-induced turbulenceto fully characterize the turbulence environment. The latter cannot, of course, be considered here In this study, TI is estimated from the TKE using the formulation

$$\mathrm{TI} = 100 \quad \frac{\sqrt{2 \quad \mathrm{TKE}}}{\mathrm{V_h}} \quad \text{in} \quad [\%] \quad , \tag{10}$$

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: $TI = \sigma_u/U$, 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-minute averages. Wind energy industry standards typically require 10-minute 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.

a1, b1: Comparison of diagnosed model TI (total and SGS) between D2 simulations at 77 m height and SKM simulations for DOM04 at 87.5 m with DL-derived TI at 95 m height at the Falkenberg observation site. a2, b2: Histograms of relative errors for total TI between ICON and Doppler lidar. Vertical dashed lines indicate a relative error of $\pm 30\%$.

Figure 8 a1 Figure 8 shows a comparison of TI between ICON and DL observations. 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 june 8th, but they reflect mesoscale variability rather than true turbulence. Focusing on the D2 configuration, the According to Eq. (10) the peak is driven by two coinciding factors: the very low horizontal wind speed ($v_h \approx 1$ m/s) 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. 8 [a]). 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. 8 [a2b]). This model error is comparable with that for the ICON wind speed simulations (Fig. 2 a2[a]) but is smaller than the TKE relative error (Fig. 2[b]). The latter implies that TI, as a normalized turbulence variable, is more robust with respect to systematic errors in wind speed and TKE. The TI from ICON at SKM scales (here 78 m for DOM04) does not necessarily improve the results (Fig. 8 b1, b2). Compared to the D2 setup the histograms of the relative errors for 1DCONF rather reveal a redistribution of relative errors towards more frequent and stronger TI underestimation with a relative error $<$ - 30 %. In contrast, for 3DCONF the histogram of the rel. errors indicates a more frequent TI overestimation with a rel. error $>$ 30 %. As the SGS-TKE fraction decreases with smaller mesh sizes (Sect.**??**), so does the SGS-TI (dashed lines in Fig.8 a1, b1) . Therefore, to provide total TI for subsequent

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 applicationsfrom higher-resolution NWP models. In order to balance accuracy with the amount of effort required, it would be necessary to run model diagnostics of the GS-TKE from intermediate prognostic wind field variables during runtime.

For applications such as evaluating the suitability of wind farm sites based on ambient turbulence, an NWP error rate of $\pm 30\%$ is high in comparison to the greater precision of more advanced models (e. g. LES). However, such simulations are more computationally expensive than NWP models so that depending on the application and risk tolerance, this error could be an acceptable compromise. Future work 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. If good agreement between the turbulence intensity 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 observed turbulence intensity can be demonstrated TI observed from DL measurements across the entire model domain, it would make sense to provide the wind energy sector with ICON-based TI products in addition to the existing data .

Regarding applications for reliable day-ahead and intraday

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 forecasts that take into account both the wind and the TI, errors in wind speed associated with a poor representation of nocturnal LLJs in the NWP model (Fig. 2 a) are currently too large to realize any additional benefits from including NWP-derived TI. This can be verified by referring to the graphs shown in Fig. **??**. Estimated wind power production (WPP) (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' is calculated based on a parametric model for of the power curve of wind turbines (Saint-Drenan et al., 2020), for two cases: i) using wind speed only, and wind speed plus ; and ii) using wind speed and TI from DL and ICON (D2 configuration) as input variables. WPP was calculated 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. Since 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 (Figs. 9 [a], [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. However, in the convective boundary layer, where wind speed errors tend to be smaller, the effect of TI on enhancing wind power production becomes apparent when estimates are calculated using observed wind speed and TI as the input data. This effect can also be reproduced using simulated winds and TI based on the NWP model. (Figs. 9 [a], [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

In this paper we demonstrate the usefulness of Doppler lidar observations for the evaluation of 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 used by the German Weather Service. 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 parameterizations in NWP models with sub-kilometer resolution by discussing two configurations of the ICON model with horizontal mesh sizes in a range from 2.1 km to 78 m. The configurations differ in their turbulence and surface transfer schemes. The first model configuration (1DCONF), which is currently used for regional NWP at a horizontal resolution of 2.1 km and was recently introduced at 500 m, is based on the "Turbdiff " turbulence schemetogether with the "Turbtrans" surface transfer scheme, while the other configuration (3DCONF) uses a 3D Smagorinsky scheme commonly used for high resolution LES, together with a Louis surface scheme. 3D winds and turbulence observations are retrieved from a conically scanning DL 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 for evaluating turbulence parameterizations under real-world conditions. As the DL measurements were obtained at the Lindenberg Meteorological Observatory in Eastern Germany using a method that provides a self-consistent data set on wind, TKE, EDR and turbulent length scale observations for eastern Germany within the lowest 600 m of the atmospheric boundary layer (ABL) , all relevant for turbulence parameterization and considered in the model evaluation. The over a five-day summer perioddiscussed is , the results obtained in this study are specific to this location and the atmospheric conditions encountered, which were characterized by a convectively mixed layer during the day and low-level jets stable atmospheric conditions that favor the development of low level jets (LLJs) at night.

Multiple lines of analysis within our model evaluation including timeseries, profile statistics and singular value spectra analysis indicate remarkably similar wind speed simulations not only by both model configurations, but also across all mesh sizes. Winds in the daytime convective boundary layer are generally wellreproduced. At night, deficiencies in the representations of the low level jet become evident. Neither configuration can substantially improve on the nighttime deficits with increasing resolution. 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. It is likely The similarities in the observed model behavior suggest that some of these conclusions may also apply to ICON, which also uses a minimum diffusivity in its turbulent schemes.

When considering TKE and EDR, clear differences between the configurations are evident. The 1DCONF is surprisingly accurate at predicting turbulent properties, regardless of mesh size. This is likely the result of including local horizontal shear terms for TKE production, additional scale-interaction terms originating from non-turbulent subgrid flows, together with a resolution-dependent turbulent length-scale in the operational "Turbdiff" scheme.This extends the utility of the scheme to resolutions that 1D TKE schemes were not originally designed for.

The deficits in the turbulence properties of 3DCONF are generally larger than those of 1DCONF, but they diminish as the mesh size decreases, being least pronounced in the convective ABL. At coarser mesh sizes, however, 3DCONF deficits also contribute to a strong overestimation of turbulence even in the convective ABL which is consistent with the findings in Doubrawa and Muñoz-Esparza (2020). During non-convective periods the poor model performance using 3DCONF suggests that the underlying Smagorinsky-Lilly parameterization with the current parameter tuning is not suitable for stable regimes, where turbulence is weak, intermittent, and highly anisotropic. Since the relevant tuning parameters, such as the Smagorinsky constant and the Prandtl number, have been assumed to be fixed throughout the day and under all stability conditionsin the model runs, future work should investigate the effect of their dynamic adaptation depending on mean shear and stratification, as proposed in Basu and Porté-Agel (2006). However, given that the two configurations use different surface schemes, we cannot rule out that some of the differences, particularly 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, stem from the surface transfer schemes rather than turbulence parameterizations.

A more detailed study of the grid-scale (GS) and subgrid-scale (SGS) contributions The relative contributions from GS and SGS to the total TKE shows that neither modelconfiguration is able to fully maintain a resolution-invariant total TKE by adaptively repartitioning the total TKE between GS- and 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 2.1km, SGS-TKE contributions. While both configurations show an increase in 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 and a decrease in SGS-TKE at smaller mesh sizes, the sum of the two contributions decreases overall for smaller mesh sizes. While this is a relatively small issue in the 1DCONF, it highlights that there is room for further improvement in the scale interactions between parameterized and resolved processes in both configurations. One known weakness of the current implementation is that only the vertical flux divergence from turbulent processes is included in the first-order momentum equations. Including the horizontal flux divergence terms as well might improve the interaction between the turbulence model and wind fields, particularly at finer mesh sizes. 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.

After validating the primary turbulence variables (EDR and TKE), the analysis was extended to include the 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 largely absent 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 this is primarily due the behavior of L_V^* is directly related to the definition of the length scales used to describe the size of the most energy-carrying SGS eddies in the model (l_S, l_B). These solely depend 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 height in both parameterizationsdistance to the surface, which are held constant during the simulation period. While the model L_V^* shows some temporal variability at smaller mesh sizes, this is due to the parameterized SGS-TKE (and therefore l_S, l_B) becoming less important relative to the GS-TKE, and does not fundamentally solve the above issue. A first 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 formulationanalysis 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.

Following the evaluation of the turbulence variables, process-oriented analyses were conducted as part of a representative case study. Over the five-day period, two nights stand out with excessive ABL mixing contributing to momentum fluxes which distort the correct evolution of LLJs compared to measured DL wind profiles . This indicates a tight coupling between errors in the turbulence parameterization and the simulated LLJ development. On these nights, where EDR values are strongly overestimated, the stability functions are found to be not effective enough to suppress turbulence production under stable conditions. 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.

To demonstrate the relevance of our model evaluation, the performance of the modelin terms of the "turbulence intensity" Another application analyzed was the model's performance in providing ambient 'turbulence intensity' (TI; a parameter combining wind and TKE information) was analyzed in the context of subsequent applicationsin the wind energy sector. It has been shown that the modeled 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 between around $\pm$ 30 %- 50 %. Exploring the impact of TI on simulated single-turbine wind energy production shows that model errors in night-time LLJ winds alone can lead to relative errors in simulated wind energy yield of up to 100 % or slightly more. Therefore, in our fictitious example, no potential benefit of TKE from ICON-NWP can be derived for performance forecasting, making its use impractical for current forecasting, particularly in situations where LLJ is incorrectly predicted. This motivates the need for further research to improve the accuracy of night-time wind predictions in the ABL. Future work needs to clarify . 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 from ICON-NWP in other regimes than LLJ can be a valuable parameter for identifying potential wind energy sites in early assessment phases .

Another application of wind and turbulence forecastingis determining the height of the mixing layer, for example for dispersion modeling. The DL provides a method to retrieve MLH as 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 height where the DL-retrieved EDR drops below a critical threshold value. The same method applied to the model EDR yields a night-time MLH more comparable to observations than a purely Richardson-number based diagnostic more typically used in NWP. The morning growth of the MLH occurs generally a bit earlier using the EDR-based method compared to the Richardson number-based diagnostics. This turbulence-based MLH diagnostic therefore has the potential to provide more frequent and accurate MLH for the stable boundary layer and during the rapidly changing growth and decay phases of the convective ABLprediction of ambient TI.

We 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 in 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.

Code and data availability. The source code for ICON is available at https://www.icon-model.org/ 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).

Appendix A: Calculation of GS and SGS TKE for LES simulations

For a meaningful comparison of simulated TKE with the observations, both subgrid-scale (SGS) and resolved, or grid-scale (GS) TKE from the SKM simulations should be considered, representative of the temporal and spatial scales corresponding to those of the retrieval. Given the Doppler lidar's elevation angle of 35.3°, the diameter of the scan circle changes with height between about 300 m at 100 m height and 1600 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. Grid-scale TKE is calculated for the SKM simulation winds saved at 5 min intervals. In a first step, the winds are interpolated from the ICON native grid to a regular latitude-longitude grid with 0.001° resolution (i.e. similar mesh size) using the CDO conservative remapping tool. Then winds are horizontally averaged over a neighbourhood $(\Delta x)^2$ for each grid point, and temporally averaged over a 30 min window. (Δx) varies vertically according to the DL scan diameter up to a height of 600 m (and is limited to a maximum of 1875 m for layers above). Wind perturbations u', v' and w' are then derived relative to these temporally-spatially averaged winds in order to calculate GS-TKE. SGS-TKE is archived at 5 min intervals and averaged to get 30 min SGS-TKE values:

$$TKE_{\text{SGS},30min} = \overline{TKE_{\text{SGS},5min}}$$

It should be noted, however, that simply averaging SGS-TKE values from multiple 5 min sub-intervals does not reproduce the true SGS-TKE of the full 30 min interval, because SGS-TKE cannot be combined by a simple mean. However, since the required detailed statistics to perform a proper variance decomposition were not available from the model run, the simple arithmetic average of the 5-minute SGS-TKE values was used. The resulting 30-minute SGS-TKE values should therefore be seen as a first-order estimate systematically underestimating the true SGS-TKE rather than an exact value.

We have also tested archiving winds and SGS-TKE at 1min intervals, but the additional temporal resolution adds little to the total TKE. We therefore conclude that most of the temporal variablity is adequately captured at 5 min intervals. The sum of this GS-TKE and the SGS-TKE predicted by the Turbdiff scheme are then added to compare to the observations. Since the Smagorinsky scheme does not predict SGS-TKE directly, we derive the SGS-TKE from the other parameters predicted by the scheme, as described in Section 2.2.

Five-day composite of simulated TKE time-height cross sections. Left block: SGS contribution to TKE (1st column) including contribution from temporal variability over 30min, and (2nd column) snapshot within 30min window. Right block: GS contributions to TKE, calculated (3rd column) considering temporal variability of winds and vertically changing spatial scale of DL cone, (4th column) considering vertically changing spatial scale, neglecting temporal variability of winds, (5th column) temporal variability of winds only (no contribution from spatial variability) and (6th column) assuming vertically invariant spatial scale, neglecting temporal variability of winds. The four rows show results for the four model domains and associated resolutions.

Five-day composite of total TKE time-height cross sections, comparing (1st column) DL retrieval and model estimates of total (SGS+GS) TKE. Columns 2 through 5 show the total TKE calculated as the sum of the 3min SGS TKE plus the four different GS TKE approximations shown in Fig. **??**.

Archiving winds and SGS-TKE at frequent intervals requires additional storage space, and operational model output is more commonly archived less frequently. How well can the total TKE be approximated if we use model output at 30 min intervals, and select at constant (not height-dependent) spatial scale to average over instead? Figure **??** shows in the left two columns the five-day composite of the model SGS-TKE from the 1DCONF across the four SKM resolutions (rows). The left-most column is the SGS-TKE estimate including the temporal variability across 30min, while the second column shows merely a snapshot of the SGS-TKE at the temporal mid-point of the 30min period. Differences are subtle and largely confined to the daytime where rapid changes (particularly those associated with cold pool passages) contribute to higher SGS-TKE values when the temporal variability is accounted for. The right side of Fig. **??** shows the five-day composite of the GS-TKE contributions calculated using four different assumptions, for all four domain resolutions. The third column shows the method described above, taking into account both the varying spatial scales and wind temporal variability at 5 min frequency, i.e. our "best estimate". For the fourth column, only the variable spatial scales are considered, but temporal variability is neglected. In the fifth column spatial variability is ignored, while in the sixth column a uniform-with-height spatial scale (roughly 1875 m) is assumed, neglecting temporal variability. It is evident that the temporal variability of the resolved winds provides an important contribution to the GS-TKE, particularly at night and at coarser resolutions. It is also evident that using the variable spatial scale, most of the GS-TKE contribution near the surface comes from temporal variability.

Figure **??** then shows the five-day composite of the observed DL TKE in the first column (reproduced on each row for easier viewing) alongside the total TKE from the model, consisting of the SGS-TKE (calculated accounting for temporal variability within the 30 min window) plus the GS contributions calculated with the four assumptions discussed in the previous figure. At all resolutions, it is beneficial to account for the correct spatial and temporal scales in the calculation of the total TKE. Accounting for temporal variability is necessary to reach the observed, higher TKE values during daytime. However, when neglecting temporal variability, the assumption of a uniform spatial scale with height does compensate to a degree for the low GS-TKE contribution near the surface that one would get using the variable-with-height spatial scale.

fig01.png

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.

fig02.png

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 $\pm$ 30 $\%$ range. Displayed USA-1 sonic (Metek GmbH) data from a 99 m tall mast 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.

fig03.png

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 (D2 configuration) and DL based observations based on the same 5-day 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.

fig04.png

Figure 4. Boxplots Time-height cross sections of SGS ([a1a] - GS- and [a4b])SGS-TKE contributions, GS ([b1d] - fraction of GS-TKE relative to total TKE and [b4e]) and total (GS+SGS) (wind speed. Panels [c1] - through [c4]) simulated TKE, and relative fraction show profiles of SGS-TKE to total TKE median (d1- d4dashed) for different mesh sizes GS, (DOM01 - DOM04dotted) SGS and model configurations (1DCONF, 3DCONFsolid) , stratified by diurnal total TKE composited for the four phases (morning transition (MT), daytime, evening transition (ET), nighttime. The boxplots are based on simulated data at 87.5 m height over of the 5-day analysis period shown day, as in FigFigure 3.?? b.

fig05.png

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]- f: time-height cross sections for ICON derived L_V^* using calculated based on Eq. (8)for different , [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 resolution conditions do not satisfy $g < 1$ in Eq. (DOM01,DOM048)and parameterization schemes (1DCONF,3DCONF).

Temporal low level jet evolution in terms of profiles every 30 min but averaged over 5 min sub-intervals for wind speed and eddy dissipation rate on 9 June 2023 (night B) for the SKM simulations (DOM04) with model configurations 1DCONF and 3DCONF (c-f) in comparison with Doppler lidar based observations (a-b).

fig06.png

Figure 6. [a]-[d]: A comparison of the temporal evolution of 30-minute 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 june 9th. [e]: Relationship between vertical wind shear (S^2) and (EDR) to assess turbulence production efficiency.

fig07.png

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 $\mathrm{Ri}_{\mathrm{crit}} = 0.2$. [b]: Simulated time-height cross section of EDR, with a white contour marking the MLH based on the threshold approach using $\mathrm{EDR}_{\mathrm{crit}} = 10^{-4}\mathrm{m}^2\mathrm{s}^{-3}$ (see Appendix A). [c]: Direct MLH comparison from the different threshold approaches against observations based on both radiosonde soundings and Doppler Lidar derived EDR profiles.

fig08.png

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 $\pm 30\%$.

fig09.png

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 (D2 configuration) 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.

Appendix A: Time-height cross sections Evaluation of wind speed and turbulence parameters for DOM01-DOM03DL-derived MLH estimated from EDR profiles

Time-height plots for wind and turbulence variables in analogy to Fig. ?? but for grid resolution DOM01.

Time-height plots for wind and turbulence variables in analogy to Fig. ?? but for grid resolution DOM02

Time-height plots for wind and turbulence variables in analogy to Fig. ?? but for grid resolution DOM03

Appendix B: Validation 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). In particular, the The MLH derived from DL based DL-based EDR profiles (hereafter DL_MLH_EDRDL_MLH_EDR) is obtained by seeking for that height at which the EDR falls below a threshold of $\mathrm{Thr_{EDR}} = 10^{-4}\ \mathrm{m^2\ s^{-3}} \mathrm{EDR_{crit}} = 10^{-4}\mathrm{m^2 s^{-3}}$. 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 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. A1 [a] indicate that at night and during the morning transitions DL_MLH_EDR 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 gives provides evidence for reliable MLH observations from DL based 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 quiet large. In such cases best agreement occurs between DL_MLH_EDR and RS_MLH_DTDHDL_MLH_EDR and RS_MLH_DTDH. Note that DL_MLH_EDR DL_MLH_EDR values above 600 m, i.e., for convective boundary layers with a depth exceeding this value could not 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 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). The comparisons in Fig. A1 show that despite the different measurement principles, comparable results can be achieved for the MLHOverall, 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.

figA1.png

Figure A1. [a]: Comparison of the mixing layer height (MLH) obtained from Doppler lidar (DL) and radiosonde soundings (RS) for the 5-day time period 8-12 June 2023 at the MOL-RAO site. DL177 MLH denotes the MLH obtained from DL based EDR profiles. RS_MLH_DTDHRS_MLH_DTDH, RS_MLH_DSDH RS_MLH_DSDH and RS_MLH_Ri 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.Of all the RS based approaches RS_MLH_QC denotes the most reliable .

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. 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).

Competing interests. The authors declare no competing interests.

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, and to Mirjana Sakradzija for providing the initial SKM experiment setup.

References

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

ASTER: ASTER Global Digital Elevation Model V003, , last Access: 2024-04-08, 2019.

Archer, C. L.: Brief communication: A note on the variance of wind speed and turbulence intensity, Wind Energy Science, 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 978-1608076673, 2013.

Banakh, V. A. and Smalikho, I. N.: Lidar Estimates of the Anisotropy of Wind Turbulence in a Stable Atmospheric Boundary Layer, Remote SensingSens., 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, Atmospheric Measurement Techniques, 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, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 922, https://doi.org/10.1029/2021MS002922, 2022.

Basu, S. and Porté-Agel, F.: Large-Eddy Simulation of Stably Stratified Atmospheric Boundary Layer Turbulence: A Scale-Dependent Dynamic Modeling Approach, J. Atmos. Sci., 63, 2074 – 2091, , 2006.

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, Quart. J. Roy. Meteorol. 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, 335–350335, https://doi.org/10.1127/0941-2948/2011/0050, 2011.

de Roode, S. R., Jonker, H. J., van de Wiel, B. J., Vertregt, V., and Perrin, V.: A diagnosis of excessive mixing in Smagorinsky subfilter-scale turbulent kinetic energy models, J. Atmos. Sci., 74, 1495–1511, , 2017.

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

Dipankar, A., Stevens, B., Heinze, R., Moseley, C., Zängl, G., Giorgetta, M., and Brdar, S.: Large eddy simulation using the general circulation model ICON, Journal of Advances in Modeling Earth Systems, 7, 963–986, , 2015.

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 Nonhydrostatic 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, 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.

Dvorský, J., Snasel, V., and Voženílek, V.: On maps comparison methods, pp. 557–562, ISBN 978-1-4244-7817-0, , 2010.

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: Chemical and Physical MeteorologyChem. and Phys. Meteor., https://doi.org/10.1080/16000870.2020.1764307, 2020.

Glazer, R. H., Grimmond, S., Blunn, L., Fenner, D., Lean, H., Christen, A., Morrison, W., and Looschelders, D.: Hectometric-scale modelling of the urban mixed layer evaluated with a dense LiDAR-ceilometer network, Weather and Climate Dynamics, preprint, 1–26, 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, Quart. J. Roy. Meteorol. 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, Boundary-Layer Meteorology, 168, 1–27, 2018.

Harvey, N. J., Hogan, R. J., and Dacre, H. F.: A method to diagnose boundary-layer type using Doppler lidar, Quart. J. Roy. Meteorol. 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, Journal of Advances in Modeling Earth Systems, 11, 1153–1188, https://doi.org/10.1029/2018MS001532, 2019.

Heinze, R., Dipankar, A., Henken, C. C., Moseley, C., Sourdeval, O., Trömel, S., Xie, X., Adamidis, P., Ament, F., Baars, H., Barthlott, C., Behrendt, A., Blahak, U., Bley, S., Brdar, S., Brueck, M., Crewell, S., Deneke, H., Di Girolamo, P., Evaristo, R., Fischer, J., Frank, C., Friederichs, P., Göcke, T., Gorges, K., Hande, L., Hanke, M., Hansen, A., Hege, H.-C., Hoose, C., Jahns, T., Kalthoff, N., Klocke, D., Kneifel, S., Knippertz, P., Kuhn, A., van Laar, T., Macke, A., Maurer, V., Mayer, B., Meyer, C. I., Muppa, S. K., Neggers, R. A. J., Orlandi, E., Pantillon, F., Pospichal, B., Röber, N., Scheck, L., Seifert, A., Seifert, P., Senf, F., Siligam, P., Simmer, C., Steinke, S., Stevens, B., Wapler, K., Weniger, M., Wulfmeyer, V., Zängl, G., Zhang, D., and Quaas, J.: Large-eddy simulations over Germany using ICON: a comprehensive evaluation, Quart. J. Roy. Meteorol. Soc., 143, 69–100, , 2017.

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, Bull. Amer. 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, Journal of Geophysical Research: Atmospheres, 125, e2019JD030 317, https://doi.org/10.1029/2019JD030317, 2020.

Hope, A. P., Lopez-Coto, I., Hajny, K., Tomlin, J. M., Kaeser, R., Stirm, B., Karion, A., and Shepson, P. B.:Analyzing "Gray Zone" Turbulent Kinetic Energy Predictions in the Boundary Layer from Three WRF PBL Schemes over New York City and Comparison with Aircraft Measurements, Journal of Applied Meteorology and Climatology, 63, 125

International Electrotechnical Commission: Wind energy generation systems – 142, , 2024. 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., Bouchet, V., Brook, J., Brunet, D., Burrows, W., Charland, J.-P., Dehghan, A., Driedger, N., et al.: The environment Canada pan and parapan American science showcase project, Bull. Amer. 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, Quart. J. Roy. Meteorol. 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. Technol., 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., et al.: Aerosol, clouds and trace gases research infrastructure (ACTRIS): The European research infrastructure supporting atmospheric science, Bull. Amer. Meteorol. Soc., 105, E1098–E1136, https://doi.org/10.1175/BAMS-D-23-0064.1, 2024.

Li, D.: Turbulent Prandtl number in the atmospheric boundary layer - where are we now?, Atmospheric Research, 216, 86–105, , 2019.

Lilly, D. K.: On the numerical simulation of buoyant convection, Tellus, 14, 148–172, , 1962.

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.

Louis, J.-F.: A parametric model of vertical eddy fluxes in the atmosphere, Bound.-Lay. Meteorol., 17, 187–202, , 1979.

Machulskaya, E. and Mironov, D.: Implementation of TKE-scalar variance mixing scheme into COSMO, COSMO newsletter, pp. 25–33, https://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter13/cnl13_03.pdf, 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, Journal of Geophysical Research: Atmospheres, 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, Experiments in 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. and 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. and 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. Technol., 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, Atmospheric Measurement Techniques Discussions, 2023, 1–53, https://doi.org/10.5194/amt-2023-153, 2023.

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, 2001.

Raschendorfer, M. and Barleben, A.: Vorhersage fluggefährdender Turbulenz und ihre Registrierung, Accessed: 2025-09-26, 2014.

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.

Reinert, D., Prill, F., Frank, H., Denhard,

Richardson, L. F.: Weather prediction by numerical process, University Press, 1922.

Rohanizadegan, M., Baldauf, Petrone, R. M., Schraff, C., Gebhardt, C., Marsigli, C., Förstner, J. , Zängl, G. , Schlemmer, L., Blahak, U., and Welzbacher, CPomeroy, J. W., and Kosovic, B.: Analysis of turbulence and turbulence kinetic energy dynamics in complex terrain, J. Geophys. Res.: Atmos., 130, e2023JD040 558, https://doi.org/10.1029/2023JD040558, 2025.

Rotta, J.: DWD Database Reference for the Global and Regional ICON and ICON-EPS Forecasting System, Version 2.5.1Statistische Theorie nichthomogener Turbulenz, Zeitschrift fü, , 2025. r Physik, 129, 547–572, https://doi.org/10.1007/BF01330059, 1951.

Richardson, L. F.: Weather prediction by numerical process, University Press, 1922.

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, Renewable Energy, 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, Quart. J. Roy. Meteorol. Soc., p. 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?, Journal of Advances in Modeling Earth Systems, 5, 117–133, https://doi.org/10.1002/jame.20013, 2013.

Schulz, J.-P. and Vogel, G.: Improving the Processes in the Land Surface Scheme TERRA: Bare Soil Evaporation and Skin Temperature, Atmosphere, 11, , 2020.

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.

Smagorinsky, J.: General circulation experiments with the primitive equations: I. The basic experiment, Mon. Weather Rev., 91, 99–164, , 1963.

Smalikho, I. N. and Banakh, V. A.: Measurements of wind turbulence parameters by a conically scanning coherent Doppler lidar in the atmospheric boundary layer, Atmospheric Measurement Techniques, 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, ocean and Atmosphere, 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 Energy Science, 8, 433–448, https://doi.org/10.5194/wes-8-433-2023, 2023.

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, Atmospheric Measurement TechniquesAtmos. 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, Quart. J. Roy. Meteorol. 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.