Response to Reviewers

May 13th 2026

We would like to thank both reviewers for their helpful comments and suggestions. Based on these, and given the length of the manuscript, we have decided to split the manuscript into two parts, with the second part to be submitted separately, as a new manuscript. For the current manuscript, we want to place our focus on the *evaluation method* for turbulence schemes based on Doppler lidar observations and what benefit and insights the method can provide for various applications, rather than on the turbulence schemes themselves. These are not novel, as pointed out by Reviewer 2, but the evaluation method is, and could be applied to any TKE turbulence scheme. We therefore retain only the discussion relating to the evaluation of the operational ICON D2 scheme, and move the comparison between Turbdiff and Smagorinsky schemes, and of the model behaviour across scales to the second part, to be (re-)submitted as a separate manuscript. This will allow us to properly address the concerns raised by Reviewer 1 regarding the different surface transfer schemes used by Turbdiff and Smagorinsky without further extending the length of the current manuscript. We hope the reviewers will find this new version more concise with a clearer focus on the evaluation method.

We also decided to re-run the model simulations with the most recent ICON release (tag 2026.04) to address concerns raised by Reviewer 1 regarding the calculation of the grid-scale TKE contributions, since a new TKE diagnostic is included in this release. Of course, this more recent binary includes a number of other model changes, however the main discussion points of the turbulence evaluation remain the same.

Please find below the response to individual reviewer comments in blue.

Reviewer 1

This manuscript presents a detail-rich model validation study with a focus on wind structure and turbulence kinetic energy (TKE) with LIDAR observations. At a measurement site in Northeastern Germany, the authors employ the ICON model with several horizontal grid spacings (while a special focus is laid on ICON-D2, dx=2km and higher-resolution research setups in the hectometric range). Furthermore, two turbulence parameterizations are employed for all resolutions: the Turbdiff scheme (roughly 1D, progrnosic TKE, based on Mellor-Yamada) and a 3D scheme based on the classical Smagorinsky closure. For validation, several innovative methods are employed, e.g., singular value decomposition, comparing cross-sections, integral length scales, etc. The manuscript is very long (more than 30 pages) and rich in detail, while its purpose is not yet entirely clear. Right now, it rather reads like an internal "state of the art" report for the ICON modelling community, but it would profit from a more clear formulation of research questions, and a clear motivation statement for the methods used. Furthermore, I am very concerned on the "comparability" of the schemes themselves, since they are coupled to two different surface transfer schemes. Some passages are difficult to understand for readers who are not familiar with ICON's turbulence schemes. Most of my current concerns are on the methods chosen, and I think the authors have to analyse more closely the dependence of the turbulence schemes on the sensible heat flux and put the results into context to this challenge. Therefore, I have not yet provided detailed comments to the results themselves. I suggest major revisions for this manuscript.

Methods and their description

I have several concerns on the methods used in this manuscript, their description is not easy to follow, some of the comparison methods seem to be unfair towards the 3D scheme, and the averaging methods to calcuate the resolved TKE are questionable (at least based on the information provided):

The two turbulence schemes and their behaviour are the centerpiece of this manuscript. However, their description in the methods part of the manuscript is severely lacking.

As a suggestion, I would first describe turbdiff scheme in a subsection, and then the Smagorinsky scheme, and avoid the "jumping" inbetween schemes in the description.

Given that the comparison with the Smagorinsky scheme is no longer part of the updated manuscript, we no longer jump between schemes. Section 2.2 has been completely rewritten to give a more comprehensive, clearer description of the Turbdiff scheme.

Furthermore, the description of the turbdiff scheme in lines 135-150 is extremely difficult to follow. The reader does not know the equations of the scheme by heart. Therefore, I would strongly suggest to add the governing equations of this scheme to the manuscript (or to the appendix).

We have now included the TKE equation (new equation 2) in our description of the turbulence scheme, and extended the references that describe individual aspects of the scheme in more detail.

Questions which pop up during reading this passage are, for example:

- what is a "classical 1D scheme with 2.5 order closure"? What does this mean? Does this stem from the classical turbulence closure assumptions or from the Mellor-Yamada framework?

  Yes, we are referring to the classification of closure assumptions based on Mellor Yamada here. We have changed the wording in the text to reflect this.

- How are the horizontal shear terms implemented? Similar to Goger et al. (2018), or with other horizontal length scales? As far as I remember, the implementation was tricky on the icosahedral grid.

  Yes, the horizontal shear terms are essentially the same implementation as described in Goger et al. 2018, and also Goecke and Machulskaya 2021. We have referenced both papers accordingly in the model description section.

- Which "scale transfer terms" have been implemented for scale-awareness of the scheme? Is there a reference?

  The new section 2.2 now fully explains the scale-transfer terms, including the appropriate references. Currently implemented is the energy transfer generated by gravity waves. A corresponding source term from convection has been developed, but is currently not used operationally.

- How is "the turbulence length scale" mesh-size dependent? Horizontal or vertical? Do you mean the vertical length scale after Blackadar? (as in Buzzi et al., 2011 Goger et al., 2018)?

  The horizontal mesh size enters the formulation of the Blackadar length scale l_B through the parameter l_m (see equation 3 and following descriptive text).

- Please elaborate of "turbulence generated by organized motions" versus "true isotropic turbulence". Why is this necessary? And how do you determine L_P_max? Is this a tuning parameter?

  This distinction is part of the scale-separation concept introduced by Matthias Raschendorfer. It accounts for the fact that classical turbulence closure assumptions are based on the premise of e.g. isotropy. However, not all unresolved motions contributing to TKE fall within this strict definition. Larger-scale (yet unresolved) circulations ("NTCs" in the manuscript) also contribute, but need different closure assumptions. Turbdiff aims to separate these contributions and apply suitable closures for each. The description of the Turbdiff scheme has been revised to communicate this concept more clearly.

  The value of 500m for L_P_max is a parameter choice. For coarse mesh sizes it represents the assumed upper limit to which isotropic circulations may grow in the above-mentioned scale-separation concept.

Different surface coupling schemes for turbdiff and 3D Smagorinsky

Given the current data, it is not possible to compare the schemes in a "fair" way, because they are coupled to different surface coupling schemes, which strongly influence the surface sensible heat flux, and henceforth one of the major forcings of turbulence generation (buoyancy production). Therefore, it is challenging to attribute any deficiency in the turbulence properties to the turbulence scheme itself within the comparison of the two schemes. Therefore, I think it is impossible to say which scheme performs better or worse, because many issues, which are visible in the turbulence schemes (e.g., overstimated TKE), might actually be present due to the different surface transfer schemes. Previous studies suggest that the turbtrans scheme performs better during convective days than the classical (and much more less complex) Louis scheme (e.g., Goger and Dipankar, 2024). I do not think that this a reason to reject the manuscript, though. However, I would strongly suggest to the authors that they address the different sensible heat fluxes and the schemes in an additional analysis, for example with a correlation plot of the dependence of the relevant turbulence parameters on the surface sensible heat flux (possibly also with observations of the sensible heat flux, I assume you have them from the EC tower). Otherwise, the entire comparison between the schemes is inherently unfair, because the turbulence schemes do not have the same "starting point" (i.e., sensible heat flux input.) In general, a fair comparison between the two schemes would only work with prescribed surface fluxes - or in a setting where both schemes are coupled to the same surface transfer scheme.

We acknowledge that the disparity between surface schemes is a fundamental problem in the comparison of the Turbdiff/1DCONF and Smagorinsky/3DCONF configurations. ICON does not currently support running both turbulence schemes with the same surface transfer scheme, and a thorough analysis of surface fluxes would further lengthen the manuscript. We have therefore decided to remove this side-by-side comparison to be published in a separate paper where we can expand on the surface flux analysis (or find a solution for running both turbulence parameterizations with the same surface scheme).

Appendix, lines 773- 783

The description here is again difficult to follow, and my remark here is mostly to invite the authors to double-check the analysis method.

Given the new, abbreviated content of the manuscript, this section has been completely rewritten.

Please describe first, how TKE is calculated from the observations - and then - in a separate subsection - how it is calculated from the model. However, my largest concern appears at line 765:

"Grid-scale TKE is calculated for the SKM simulation winds saved at 5 min intervals." - Are these intervals already averaged temporally?

From this description, it is not entirely clear how the temporal 5-minute data are derived. Is this instantaneous model output every 5 minutes? Or are the 5-minute intervals already averaged from high-frequency model output, for example at every time step (i.e., meteogram output)? Currently, I think the only (realistic and reliable way) to obtain high-frequency output in the ICON model is via the meteogram option for selected grid points. If the 5-minute intervals indeed stem from 5-minute instantaneous output (non-averaged), which I hope is not the case, it is impossible to derive the resolved TKE (for both schemes), because turbulent motions are relevant on time scales much shorter than 5 minutes (or even one minute) - you would need model output at every time step. This would deteriorate all your findings in Section 5.1 (and their interpretation); and furthermore, this methodological flaw would be a reason for rejection of the MS, especially when you interpret the data from the sub-hectometric range.

We did initially calculate the temporal wind variance for the grid-scale TKE based on instantaneous wind output every 5min. However, since the initial submission of the paper, the authors have implemented a new diagnostic in ICON (now part of the new release tag icon-2026.04 used in the updated manuscript) which calculates the wind variances „on the fly“ from wind fields at every model time step. The description of the GS TKE calculation has been updated accordingly, and was moved from the Appendix to Section 2.3.

We did however investigate how much of an error is introduced when using 5min instantaneous winds for the calculation of the GS TKE component, and find that this is a) dependent on the model mesh size, and b) not as significant an error as one might expect. For the case described in the updated manuscript (relatively coarse 2.1km resolution), the GS TKE component is naturally much smaller than the SGS contribution, such that an error in the GS component does not contribute as strongly to the total TKE error. The GS TKE is underestimated (overestimated) approx. 57% (43%) of the time. 99% of samples fall within +/-25% of the total TKE calculated with the online diagnostic, 63% within +/-5%.

Nevertheless, the new online diagnostic is clearly the correct solution to this problem.

Further questions on the methods:

- How do you deal with changing horizontal resolution in your spatial averages?

- What is is the difference between grid-scale averages and SGS time averages? Given the description, it sounds similar.

- line 773: are the 5-min 'sub-intervals' already averaged?

Minor comments

line 24: Please cite relevant references for the turbulence gray zone, e.g. Honnert et al, 2020, and Wyngaard er tal. (2004)

The original manuscript already cited Honntert et al. 2020, but we have now also included the reference to Wyngaard 2004. L29

line 33: What is the "standard ICON turbulence scheme"?

The Turbdiff scheme was meant here, however this sentence no longer exists in the rewritten manuscript.

lines 39-46: Since the turbdiff scheme uses a prognostic TKE equations, it would be fair to mention TKE budget-based model validation methods as suggested by Goger et al, 2018, Nilsson et al (2016), and Rohanizadegan (...)?

We now also refer to budget-based TKE evaluation studies in our introduction. L38 line 82: Please describe the location and its surroundings a bit closer - not every reader knows the observatory and its location. Furthermore, where exactly is the instrument placed? After reading the entire manuscript, I would strongly suggest to add a figure showing the model domains and adding the location of the DL on the map.

We have extended the description of the measurement site L73-77. Given that we now no longer discuss simulations from the nested domains, a figure no longer seems necessary.

line 121: Here - at latest - adding the figure with the model domains would be appropriate.

See above.

line 126: "SKM horizontal resolution" - maybe I've missed it, but what does this abbrevation mean?

SKM stands for "sub-kilometer" - this abbreviation is no longer used in the updated manuscript.

line 155: "3D like extensions". This is very ambigous. What 3D effects are you talking about?

This wording is indeed ambiguous, and we no longer use it. We have endeavoured to be more specific in our description of section 2.2 more generally, and have refomulated this paragraph to read: "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."

line 162: It should be clarified that the Blackadar length scale is vertical

Done. L149

line 175: at this point (at latest), the prognostic TKE equation should be introduced.

The prognostic TKE equation is now included in section 2.2.

line 178: Could you please provide a reference on the equation to estimate EDR from the Smagorinsky scheme?

This point and the following two relating to the Smagorinsky scheme are no longer applicable to the revised manuscript.

line 176: "Underlying the Smagorinsky scheme is the assumption that turbulence

generation via shear production and buoyancy production/destruction is balanced by dissipation" - this is not specific to the Smagorinsky scheme, but a general concept in boundary-layer meteorology (Stull, 1988).

line 194: "ICON turbulence scheme" - be consistent with the naming of the schemes. Both turbdiff and 3D-Smag are "ICON turbulence schemes".

line 208: You might inducate somewhere that cold pools are associated with convective systems, i.e., thunderstorms

We have adapted our description of the general weather situation during the five day period to reflect this. L210

line 214: I am aware that this is not the focus of the MS, but to double-check, you could look at RADAR composits to make sure...

We have, in fact, looked at radar composits and see convective precipitation cells in the vicinity of the Falkenberg measurement site (though not travelling directly overhead). Thus we have reason to believe the temperature drops (together with sudden change in wind direction) observed at the Falkenberg mast are associated with cold pools. Without a distributed sensor network as used during FESSTVaL, this still remains somewhat speculative, hence the cautious wording in the text. L226

line 380-81: Could it be that these unrealistically high TKE values are related to the higher sensible heat fluxes compared to the turbdiff scheme? So the scheme itself is likely physically consistent, but you just have a slightly different surface forcing.

This part of the discussion is no longer part of the manuscript.

Reviewer 2

The manuscript evaluates the ICON weather model at sub-kilometer resolution using Doppler lidar observations. The existing turbulence parameterizations (Turbdiff and Smagorinsky) are examined to assess how well they represent small-scale turbulence parameters. Given that these parameterizations already exist, my main concerns remain regarding the novelty and motivation of this study. In addition, several major questions arise regarding the validation methodology, the relevance of Doppler lidar observations, and the acceptance criteria for both measurements and modeling, all of which indeed depend on the intended applications. This review focuses solely on the validation and experimental aspects of the study, in line with my expertise in remote sensing, and does not assess the model development or physical parameterizations in detail.

Major comments:

- The authors need to provide justification for the novelty of this study, given that the ICON model and those turbulence parametrizations (Turbdiff and Smagorinsky) already exist. Also, the main motivation should be clearly mentioned.

  As already stated in the introduction of our response to both reviewers, the primary objective of this study is to use DL-based measurements for wind and turbulence variables (e.g. TKE, EDR, integral lengths scales) for evaluating the Turbdiff turbulence parameterization used in the operational ICON configuration (D2). In particular, by employing the retrieval method developed by Smalikho and Banakh (2017), a novel kind of internally consistent dataset for wind and turbulence variables is used through the use of Doppler Lidar measurements from a single scan strategy. Unlike conventional approaches that rely on alternating scan modes, this method ensures spatio-temporal consistency by eliminating the need for sequential switching between different measurement configurations. This is particularly beneficial for model evaluation, as it enables a physically consistent assessment of the relationships between wind and turbulence characteristics without additional uncertainties introduced by differing observational sampling properties.

  In the revised manuscript, we have further emphasized this aspect of novelty within the introduction to highlight its significance more clearly.

- Page 28, line 630, Eq. 8 appears to be incorrect. TI is defined as the ratio of wind speed standard deviation to the mean wind speed. Please refer to Eq. 9 and Fig. 1(b) in the following article, which clearly explains the difference. Archer, C. L. (2025). Brief communication: A note on the variance of wind speed and turbulence intensity. Wind Energy Science, 10(7), 1433–1438. https://doi.org/10.5194/wes-10-1433-2025

  We would like to thank the reviewer for this valuable feedback. We have taken this suggestion into account and have addressed this point in the revised manuscript in lines 488 to 502.

- Page 3, line 64. The authors consider a wind lidar for validating small-scale turbulence parameters (e.g., TI), noting that it provides sufficient accuracy compared to mast measurements. This statement is misleading. In IEC standards, instruments that provide point measurements (e.g., cup anemometers) are considered reference instruments and serve as industry best practice (sufficient accuracy) for general wind energy applications. However, for the wind lidar system, there is only a recommended practice for acceptance error criteria, depending on the application (see the DNV report below). This is due to the major challenges in lidar-based TI estimation (including cross-contamination, volume averaging, stability dependence, and noise in measurement). Therefore, the main question is whether wind lidar can achieve sufficient accuracy for validating small-scale turbulence parameters (?). If so, the authors should discuss the acceptance error criteria for both measurement and simulation, depending on the use cases. For example, in Figure 14, even if the substantial uncertainty in lidar-based TI measurements is ignored, the relative error distribution remains significantly above 10%. This level of accuracy is insufficient for wind energy applications (including both load analysis and energy production assessments). So, I would recommend revising the manuscript and conducting a quantitative error analysis from an application perspective. DNV. Lidar-measured turbulence intensity for wind turbines. Recommended practice, DNV-RP-0661, Det Norske Veritas, 2023. 2

  We thank the reviewer for this important comment regarding the role of lidar measurements in wind energy applications and the reference to IEC standards. We agree that, according to IEC recommendations, point measurements such as cup or sonic anemometers are considered as reference for TI, and that DL-based TI estimates are subject to several well-known limitations, including volume averaging, cross-contamination, stability dependence and measurement noise.

  To avoid any misunderstanding, we wish to specify that the primary objective of this study is not to assess DL measurements in the context of IEC-compliant wind energy applications. Instead, the focus is on the evaluation of the Turbdiff turbulence parameterization used in the operational ICON configuration (D2).

  The references provided in the introduction (i.e. Wildmann et al. (2020)) and Sect. 2.1 (i.e. Päschke and Detring (2024)) of the revised manuscript give insights that with the applied DL data processing strategy, the DL-derived turbulence quantities show a high level of consistency with colocated sonic anemometer mounted at a 99m mast at the measurement site. The additional comparisons of DL wind and turbulence variables with sonic anemometer data further show good agreement over the five-day test period considered in our study. This provides increased confidence in the reliability of the DL measurements and supports their use for model evaluation applications.

Moreover, we would like to emphasize that: (i) the applied retrieval method includes correction terms taking the volume-averaging effects and errors due to instrumental noise into account, and (ii) in a pre-processing step of the retrieval an advanced filtering procedure was used to reduce noise contamination due to the low number of pulses used for the measurements. In Päschke and Detring (2023) it has been shown that these steps can significantly contribute to improve the quality of the DL-derived turbulence data, resulting in a high degree of agreement with sonic measurements.

Although this study primarily aimed at evaluating the turbulence parameterization of the model, the final section explores the potential applicability of the model results for selected downstream applicatons. TI, which is of particular relevance for wind energy applications, was chosen as an example. It is emphasized that the TI analyzed here represents only the ambient atmospheric turbulence conditions and not the complete turbulence environment within a wind farm, since wake-induced turbulence effects are not included. While the model bias in ambient TI is non-negligible, such dynamically simulated ambient TI fields may still offer added value compared to static or climatological reference values often applied in practical wind energy applications. A discussion of acceptance error criteria is 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 transparent and application-independent benchmark, which can be used by the wind energy community to judge the suitability of the data for their respective purposes.

We have placed greater emphasis on this aspect in Section 6.2 of the revised manuscript.

Minor comments:

- Page 1, line 1. The abstract should include numeric values to quantify the main findings.

  We appreciate the reviewer's suggestion to include quantitative error metrics in the abstract. Although statistical analyses were performed, the study mainly follows a process-oriented and descriptive evaluation across different variables and atmospheric regimes. We therefore found it difficult to summarize the results using only a small set of representative numbers without oversimplifying the findings, and decided to retain the current formulation of the abstract.

- Page 3, line 82. Please specify the correct commercial name of the instrument as well as the corresponding citation. There are several instruments with similar names (XR, XR+, VS+).

  We were working with a Halo Photonics StreamLine (Standard/Pro) system manufactured in 2019. In current catalogs, its closest equivalent is the StreamLine AllSky. However, it is technically distinct from the XR (Extended Range) series because it features a 180 ns pulse length and a variable focus adjustment, whereas the XR models typically use longer pulses and a fixed focus.

  Due to the system's manufacturing date in 2019, its technical specifications do not fully align with current product nomenclature; for this reason, we cannot provide more specific model designations as requested.

  As an alternative, we have included additional technical specifications for the LIDAR used in Section 2.1.

- Page 4, line 94. Please specify the scanning pattern. If one of the standard methods is used (e.g., velocity azimuth display), please use the standard terminology instead of "conical scan patterns".

  We thank the reviewer for this comment and agree that the term "Velocity Azimuth Display" is widely used in the literature. We would like to clarify, however, that the two terms refer to different aspects of the measurement and analysis framework."Conical scanning" describes the measurement geometry, i.e., the scanning strategy in which the lidar beam follows a cone at a fixed elevation angle. In contrast, "Velocity Azimuth Display (VAD)" refers to a specific analysis method that uses the azimuthal variation of radial velocity to retrieve wind parameters.

  In this study, we follow the approach of Smalikho and Banakh (2017), which is formulated in terms of conical scanning and extends beyond classical VAD applications by including the retrieval of turbulence quantities such as TKE and EDR. To remain consistent with the original methodology and its theoretical framework, we therefore retain the term "conical scanning", while acknowledging its conceptual relation to VAD-type methods.

- Page 4, line 99. According to the information provided, each scan lasts for 72 seconds. Therefore, the sampling rate of wind speed is 1/72 Hz and doesn't seem to be a "high temporal resolution". Also, the accumulation time (or the number of pulses) could be more relevant for SNR discussion (compared to scanning speed).

  We have removed the phrase "high temporal resolution" and added more detailed information about the number of pulses used in Section 2.1.

- Page 4, line 103. About filtering noise, is there any specific reason for selecting the filtering method developed by Päschke and Detring (e.g., poor performance of factory-set SNR threshold)? And how much has the availability changed compared to the standard factory-set filtering?

We thank the reviewer for raising these questions. Since the primary focus of the manuscript is the model-observation comparison rather than the measurement technique itself, the specific challenges related to the measurements were not discussed in detail. The requested information is discussed in detail in Päschke and Detring (2024), which was cited in the original manuscript as a source for additional methodological information.

As demonstrated in section 3 of the cited reference, the use of a standard SNR-threshold filtering approach results in the rejection of a considerable amount of valid data when the number of pulses is small. This motivated the development of an alternative filtering method, which is described in detail in the referenced study. Additionally, Fig.16 in this reference clearly shows the differences in the final data availability if either a classical SNR threshold approach or the suggested filtering procedure is used.

However, it is important to point out that this certainly applies to the systems available at our location, as we have specifically examined them for this purpose. We are not aware whether the same applies to Doppler lidars manufactured by other companies.

- Page 8, Figure 2(c). The EDR availability from Sonic is lower than that of the two other panels (a) and (b). Do you expect different data availability for Vh, TKE, and EDR in Sonic measurements? Why?

  The derivation of mean wind quantities and turbulence-related variables such as TKE and EDR relies on different underlying assumptions. Since the retrieval of turbulence quantities involves additional assumptions beyond those required for mean wind estimation, the corresponding quality-control criteria may differ substantially. In particular, the estimation of TKE and EDR relies on assumptions related to turbulence structure and inertial-range behavior. As a result, turbulence retrievals can fail quality checks under conditions where the associated wind measurements remain fully valid.

  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.

  The missing sonic data is also due to a quality control check for wake effects on the 99 m mast at the measurement site.

  We have included explanations of these differences in the caption for Fig. 2.

- Page10, Figure 3. I'd recommend checking all figures and removing the extra white rectangles near the y-axis (e.g., Fig 3, 6, 7, etc.).

  Thank you for the note. As part of the script revision, we have recreated the figures and made sure to avoid the "extra white rectangles".

- Page 16, Figure 6(c). Could you please explain the relative error formula? In some cases, the error between the model and measurement exceeds two orders of magnitude (e.g., MT 06-11), yet the histogram shows errors of only up to 400%.

  The rel. error calculation was performed in accordance with the general formula for relative error, i.e.:

  rel. error $= \frac{x_{measured} - x_{reference}}{x_{reference}} \times 100\%$

  For the sake of readability, the visualization in the histograms was restricted to errors up to 400%. Although individual cases show higher deviations upon closer inspection of the time series, these represent isolated outliers that do not reflect the general distribution characteristics. In the revised manuscript we added a note on that in the caption of Fig. 2.

- Page 28, Figure 14. What's the reason for this peak? Is it because of the low mean wind speed?

  Yes, the significant peak in TI is observed during the passage of a cold pool (see Sec. 3 in the manuscript). 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, likely caused by the high shear and instability at the gust front.

  We made an additional point on that in section 6.2 of the revised manuscript.

- Page 28, line 635. It should be noted that the standard averaging time for wind speed and TI is typically 10 minutes according to IEC standards (not 30 minutes). IEC 61400-1, Wind energy generation systems - Part 1: Design requirements, ISBN 9782832279724. 2019.

  We agree with the reviewer that the 10-minute averaging period is the common industry standard. We have revised the text (see Section 6.2) to state that 30-minute averages were used in this study and acknowledged the potential differences in turbulence characterization compared to the 10-minute IEC standard.

- In general, the manuscript is long and could be more concise to highlight its key contributions. Lastly, I appreciate the review opportunity provided by the Geoscientific Model Development (GMD) journal, and hope that these suggestions will be constructive in helping the authors enhance the overall clarity and contribution of their work.

The steps undertaken to shorten the manuscript are outlied in the introductory section of the reviewer responses.