Referee comment (RC) on

"Wind and turbulence evaluation of the ICON model (icon-2024.01-1) at sub-kilometer scales using Doppler lidar observations."

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:

1- 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.

2- 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

3- 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.

Minor comments:

1- Page 1, line 1. The abstract should include numeric values to quantify the main findings.
2- 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+).
3- 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".
4- 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).
5- 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?
6- 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 V_h, TKE, and EDR in Sonic measurements? Why?
7- 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.).
8- 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%.
9- Page 28, Figure 14. What's the reason for this peak? Is it because of the low mean wind speed?
10- 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.
11- 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.