Comments to egusphere-2026-5393
“Horizontal wind estimation from consumer-drone hover telemetry: comparing machine-learning and calibration approaches” by Polle et al.

General comments

The authors address a relevant topic with a well organized methodology and can derive results that are useful for the scientific community. Robust techniques that enable to derive the wind speed from drones’ pitch and roll can, indeed, pave the way for low-cost strategies to observe wind properties in the atmospheric boundary layer, which are extremely valuable. In general, I suggest adding a more fluid narrative to the manuscript, as in many places it seems to be a list of fragmented analyses. In this sense, I would consider merging some paragraphs and sections, making sure to connect them with smooth transitions. This would improve the readability of this nice research paper. Overall, I suggest minor revisions, as detailed in the following list of points.

Specific comments

L3 The Ridge regression is referred to as a linear model despite containing nonlinear terms involving pitch and roll, as described in Table S2. I understand that the model is linear with respect to “features derived from pitch and roll” (L38) but I find this aspect confusing. If nonlinear terms are included in the model why should it be called linear? I also suggest that the Ridge model is described in more detail. For instance, how are the variables listed in Table S2 selected? Can you define the L2 penalty?

I would consider adding a picture of the drone when it is on land, with its dimensions, to better visualize its configuration.

L54 Is the choice of the height of the reference anemometer (1.6m) enough to rule out that the circulation induced by the rotors close to the surface is not affecting the measurements?

L61 The choice of the length of the moving average (30 s) is mentioned here for the first time in text and is discussed in S1.6 - Window-length sensitivity. However, it is properly introduced in L91. Please make sure to describe the filtering before referring to it. To me, the choice of the length of the moving average deserves a longer discussion in the main article (expanding LL93-95) because it is related to the physics of the problem. To my understanding, it has to do with the inertia of the drone, which poses a limit on how fast the drone adjusts to fluctuations in the wind. You mention that the error increases for shorter time windows: have you looked into the amplitude of the fluctuation with respect to those of the reference anemometer? Could the use of a shorter averaging time window retain more information on the wind fluctuations (as the data are recorded at 10 Hz), despite corresponding to a larger mean bias? I suggest elaborating more on this point.

L71 Why is there a time lag between -1 and 4 s? Does it have to do with the inertia of the drone? What else could physically explain it and its variability? As you can measure it, why is it not corrected to quantify the RMSE and bias of the wind estimates?

Fig8 It is very clear that the goal here is to show a case with high wind speed, but I suggest including two more panels showing examples of time series with medium and lower mean wind speed.

L294 It is mentioned that no air density measurements are performed at this stage and they could improve the results. I suggest expanding this point to provide a feeling of how important this information could be, especially using Machine Learning approaches. Along this line, I think that the Conclusions section would benefit from a discussion that tries to identify the highest priority activities to be performed: e.g. testing different platforms or different heights, including air density measurements, exploring stronger wind regimes (any other large-scale atmospheric variable that can play a role?), etc.

Technical corrections

L43 Is a “fold” a “session”? Please clarify what fold means.

Fig1 The symbols (S, D, theta, etc) should be defined in the main text when the figure is presented and then used consistently in the manuscript.

L65 Please describe what the RC.rudder stick is and what it is for.

L105 What is the distance covered by the flights required for the calibration baseline?

L199 What is the added value of showing the window-level performance? As discussed, in fact, it gives more weight to the longer sessions. Why not showing Fig6 for session performances? Too few points? What about extracting the same number of windows for each session and using these windows?

L236 Please introduce the t-test you are doing with more details, as the t_{49} variable appears out of nowhere.

L280 No need to repeat the statistical performances here.