This manuscript investigates how short measurements could be long-term corrected (LTC) with reference data from ERA5 in order to obtain accurate estimates of long-term mean wind speed and energy production. The motivation is that full-scale (1y+) measurement campaigns are costly and often unfeasible for small-scale wind power projects.

Overall, the paper is clear and the study is robust. The dataset of wind speed measurements is relatively large and representative for small-scale wind power in the US. The manuscript addresses a real problem, which is well described, and is thus relevant. A nice aspect is the focus on worst case scenarios in addition to mean statistics.

There are however some possible improvements to methodology and text. Below are comments listed approximately in order of importance.

1. ERA5 is state-of-the-art for LTC and as input to mesoscale models. Due to e.g. its coarse resolution it is however not suitable for mean wind estimates at particular sites. Raw ERA5 is thus a poor baseline and all comparisons becomes more or less irrelevant. It would strengthen the paper greatly if e.g. GWA (global wind atlas, contains mean wind speed and parameters needed for energy calculations) was used as benchmark.

2. Overall, the paper is well written and methods are clear. The paragraphs on MCP and quality threshold (L197-207 + L223-235) are however difficult to follow in detail. As an example, it is not fully clear weather the authors exclude MCP-simulated wind speeds for periods with missing observations when computing means. A shorter, clearer and stepwise description would be better. From my understanding

    a. The mean of the full observation period is taken as ground truth.

    b. A model is trained on a short (few months) period

    c. The model is used to predict the full observation period.

    d. Mean statistics are computed and compared to ground truth. Samples with missing observations are excluded from the calculations.

3. One of the main reasons why MCP-methods will work well or not on short measurement series is the seasonality of the errors in the reference series. It is not a problem per se if the measurement period is not representative for the long-term as long as the reference series accurately captures the variations. As an extreme example, with perfectly correlated observation and reference series, one observation sample is enough to estimate the true long-term mean wind speed. In practice, as is well demonstrated in the manuscript, the risk of getting poor results are higher if measurements are taken during e.g. low-wind periods. But these phenomena should not be confused (as in e.g. L24-27, L440-442).

4. Energy results should really be in the results section (not in discussion). Since energy production is more important than mean wind speed for the intended application, it would be good to emphasize these results more (on the expense of mean wind speed), e.g. in discussion and abstract.

5. It would be good to include descriptions and references of state-of-the-art MCP methods.

6. Many different regression tree algorithms exist. State which one you use and give a reference.

7. The quality control of measurements is performed automatically. This is fine, in particular given the large number of observations. Periods with disturbed observations might however persist after the control, an example is partial icing of the anemometers. This could maybe explain the poorer results in Alaska.

8. L33-34: normally, one describes it as that reference data (e.g. ERA5) are used to correct a short measurement, not the other way around.

9. L58-65 could be rewritten to sound less like a sales text for ArcVera.

10. Refs to fig 5c should be 5d (L357 + L428)

11. L398 "versus performing a bias correction" could be "versus estimating mean wind speed" (the term bias correction should not be reserved for mean wind speed estimates).