General Comments

The authors address a valid challenge in wind resource and wind project design representing 50-year wind speeds with limited data. They derive return values from four different datasets, two modeled and two measured, with the goal of showing that a comparable V_{ref} can be obtained from a selection of modeled storms as longer-term continuous measurement campaign. The communication is written clearly and succinctly with an emphasis on usability. Some clarifications and refinements to help readers replicate/implement the findings are posed below.

Specific Comments

1. WES defines *Brief Communications* for both brevity and impact. Please provide context for designating these findings as unique and high-impact, with justification for its implementation over other methods. 1 year of data in itself does not present a sufficient sample size to calculate a 50-year return value.

2. P. 3: Please specify if the Weibull distribution used is 2-parameter or 3-parameter.

3. Measurements: Please specify the measurement interval (i.e., 10 mins) and any differences (i.e., due to shear) between the 80m measurement values and 100m ERA5 values.

4. P. 3: While DNV guidance may recommend Weibull for "extreme metocean parameters", ocean and atmosphere parameter behavior can differ. This distribution is widely used for wind speeds, however, please provide justification of its suitability at the investigated height for the study location (such as results of a chi squared or Kolmogorov-Smirnov test).

5. P. 4: Please describe the ERA5 highest wind speed calibration choice and method in more detail. What is the threshold to qualify as a high speed?

6. Calibration of two datasets (ERA5 and ASPIRE) appear to be derived from the same 1-year period, and that this year is maintained in the 20-year measurement EVA.

    a. How is interannual variability quantified for the site? How does the calibration treat this—what is the sensitivity to different calibration periods?
    b. Please show that the year used for calibration is representative of long-term conditions at the site. How was the year selected? From the return value estimation on P. 4, it appears that this year was particularly stormy.
    c. How does the V_{ref} for the ASPIRE and Cabauw compare when the training (calibration) year is removed from the measurement dataset (19 instead of 20 years)?

     d. In the case of the ASPIRE dataset, it seems that extreme events (45 storms) are calibrated from a continuous sample (i.e., including many normal conditions). Please clarify in the text if this is incorrect, or please justify why these storms are calibrated to the values chosen.

7. Are the 45 storms from the ASPIRE dataset are taken from 1 year? Please show that that measurements and model values, prior to POT analysis, are comparable. Please also describe the range of storms selected (i.e., from X to Y peak winds), and how storms were selected.

8. Please provide more detail on how the ASPIRE dataset is related to the climate, including details on its parent mesoscale simulation and its parameterizations.

9. The argument of the paper would be bolstered by showing similar performance when using different 1-year measurement periods and for different locations (i.e., different from a homogeneous, flat pastured area). If available, please show this.

Technical Corrections

1. P. 3: (b) implies that a subset (less than 1 year) of data from the Cabauw campaign is used for calibration. I interpret that 1 year of the Cabauw dataset was selected for calibration. If so, please update for clarity.