Review of manuscript "Estimating Long-term annual energy production of large offshore wind farm from large-eddy simulations: methods and validation with a 10-year simulation."

The authors present three different approaches to represent long-term effects (defined in the order of 10-year averages) in a hypothetical wind farm in the North Sea by using large-eddy simulations of a period smaller than that of a full calendar year. Their approach and results are interesting; however, I would only recommend their work for publication after they carefully address the following:

1) In section 2.1 the authors layout the foundation of their approach. According to the authors, their approach involves the application of Bayes' theorem to the continuous random variable P which is the power production of the entire wind farm, and the continuous random variable M which is the wind speed at a nearby location. To this end, the authors define the conditional densities in equations (1) and (2) however they do not explain how Bayes' rule is applied. What appears to be happening is the conditional densities are integrated to obtain the probability density for the wind farm power production obtained by means of LES, $f_L(P)$. This is indeed based on Bayes' theorem, but further explanation is needed. If we start from Bayes' theorem we have:

$$h_{L\,|\,ERA}(P, M) = \frac{h_{ERA|L}(M,\ \ P)\, f_L(P)}{g_{ERA}(M)},$$

where $h_{L\,|\,ERA}(P, M)$ is the conditional probability of the power predicted correctly by LES given that the ERA5 reanalysis data provides true values for the wind speed, equals to the likelihood of the wind speed being calculated correctly by ERA5 given that power is calculated correctly by LES and multiplied with the prior probability density, $f_L(P)$, and the marginal probability density, $g_{ERA}(M)$. What the authors do not mention is their main premise which is that the likelihood of the wind speed being calculated correctly by ERA5 given that power is calculated correctly by LES is equal to 1. This allows them to integrate over the wind speed M to obtain

$$\int h_{L\,|\,ERA}(P, M) g_{ERA}(M) dM = \int f_L(P)\, dM = f_L(P)$$

During integration they also use the fact that $f_L(P)$ is independent of the wind speed (and that $\int dM = 1$). To apply Bayes' theorem the authors, need to also consider that $f_L(P)$ and $g_{ERA}(M)$ to be independent probability densities, which I guess is self-evident by the fact that the two distributions have been synthesized from different datasets. This is also something the authors need to emphasize.

2) The second assumption they make is that the conditional probability between power and wind calculated for 1 year approximates the long-term counterpart. This allows the authors to calculate the long-term probability density of power $\hat{f}_L(P)$, by

only using information from the long-term wind speed probability density, $\hat{g}_{ERA}(M)$ which can be easily obtained from the ERA5 record, and the conditional probability calculated from the down-selected days, $h_{L\,|\,ERA}(P,M)$. This assumption is attempted to be validated in section 4.2, and more specifically in figure 4d, however

a. Data are shown only for scenario 1 (full-year simulation)
b. The phrase "their general shapes largely agree" in line 248-249 cannot be used instead of a quantitative metric.

My recommendation to the authors would be to use a rigorous metric such as the Perkins Skill Score (PSS) or a goodness of fit test, such as the Kolmogorov-Smirnov test, to measure how well the conditional probability densities, $h_{L\,|\,ERA}(P,M)$ and $\hat{h}_{L\,|\,ERA}(P,M)$ agree with each other. Rigorously quantifying the matching between the 1-year and long-term conditional probabilities, will provide more value to the study and increase its overall impact.

3) LES resolution may not be sufficient. While I fully agree that the authors have provided results from a state-of-the-art, meso-microscale coupled model and have therefore been pushing the limits of wind farm modeling (including resolution), the statement *"…Although this can be considered a coarse resolution, Baas et al. 2023 showed that refining to 60 m has a relatively small effect on total aerodynamic* losses of a 770 MW wind farm…" is problematic. The reasons are the following:
    a. The two studies consider different size wind farms 960MW versus 770MW and different array densities 7.2MW/km2 versus 10MW/km2. This may result in a different number of nodes used to cover the turbine spacing, so it is not really a direct comparison.
    b. Wakes remain unresolved when using either a 120m or a 60m resolution and therefore the small change in power losses should not be used justify accuracy particularly when lacking validation.
    c. The authors in Baas et al 2023, provide a much better reason for why a resolution of 120m is selected: "*This choice results from a trade-off between computational cost and accuracy and has been tested extensively in an operational setting*". I fully agree with this statement. Such studies have been pushing the state of the art of offshore wind farm modelling and they must not be judged based on previous LES studies that have only considered canonical ABL cases (or what I call turbulence in a box). I suggest the authors re-phrase this part of the paper to provide a similar statement.

Overall, the paper presents a novel an interesting approach (based on Bayes' rule) to correct for long-term effects, but the structure of the paper is not clear, and it requires additional effort by the reader. In addition, the presentation of the results needs also to be improved to allow the authors to better highlight their key findings.