We thank the reviewer for their comments on the manuscript. Although the discussion has not ended yet, we will already reply on this review here, and incorporate the second one later. The reviewer has raised three very valid points, which we plan to address in the following way:

1. We agree that our explanation of the use of Bayes' theorem (in section 2.1) is lacking. In fact, in the long-term correction method, we only use the definition of the conditional probability density, from which Bayes' theorem is a direct result. For example, for continuous random variables X and Y, the definitions of the conditional probability densities are:

$$f_{X|Y}(X,Y) = \frac{f_{X,Y}(X,Y)}{f_Y(Y)}, \text{ and } f_{Y|X}(X,Y) = \frac{f_{X,Y}(X,Y)}{f_X(X)}, \tag{1}$$

and therefore:

$$f_{X|Y}(X,Y) = \frac{f_{Y|X}(X,Y)f_X(X)}{f_Y(Y)}. \tag{2}$$

Eqn. 2 is Bayes' theorem, whereas in the manuscript, we only used eqn. 1, in the following form:

$$h_{\mathrm{L}\,|\,\mathrm{ERA}}(P,M) = \frac{h_{\mathrm{L,\,ERA}}(P,M)}{g_{\mathrm{ERA}}(M)}, \tag{3}$$

from which the remainder of the derivation follows.

Therefore, we should not have written that we use Bayes' theorem, but that we apply the definition of the conditional probability density (from which Bayes' rule is the direct result). We have adapted this in the manuscript, and added further explanation.

2. A more rigorous quantification of the match between short- and long-term conditional probability densities is indeed a good idea. We therefore added (for scenario 1) the Perkins Skill Score as a function of wind bin ($S(M)$) in the manuscript's Fig. 4:

$$S(M) = \int \min(\hat{h}_{\mathrm{L}\,|\,\mathrm{ERA}}(P,M),\ h_{\mathrm{L}\,|\,\mathrm{ERA}}(P,M))dP. \tag{4}$$

The new version of the manuscript's Fig. 4 is shown here in Fig. 1. Also, in a similar way, the error in the long-term corrected value can be split up per wind bin, thereby quantifying the final effect of the degree to which the approximation $h_{\mathrm{L}\,|\,\mathrm{ERA}}(P,M) \approx \hat{h}_{\mathrm{L}\,|\,\mathrm{ERA}}(P,M)$ holds:

$$E(M) = \int ((\hat{h}_{\mathrm{L}\,|\,\mathrm{ERA}}(P,M) - h_{\mathrm{L}\,|\,\mathrm{ERA}}(P,M))\hat{g}_{\mathrm{ERA}}(M)PdP. \tag{5}$$

Because it takes into account the mean power production value in each wind bin, and the occurrence frequency of each wind bin, $E(M)$ integrates to the

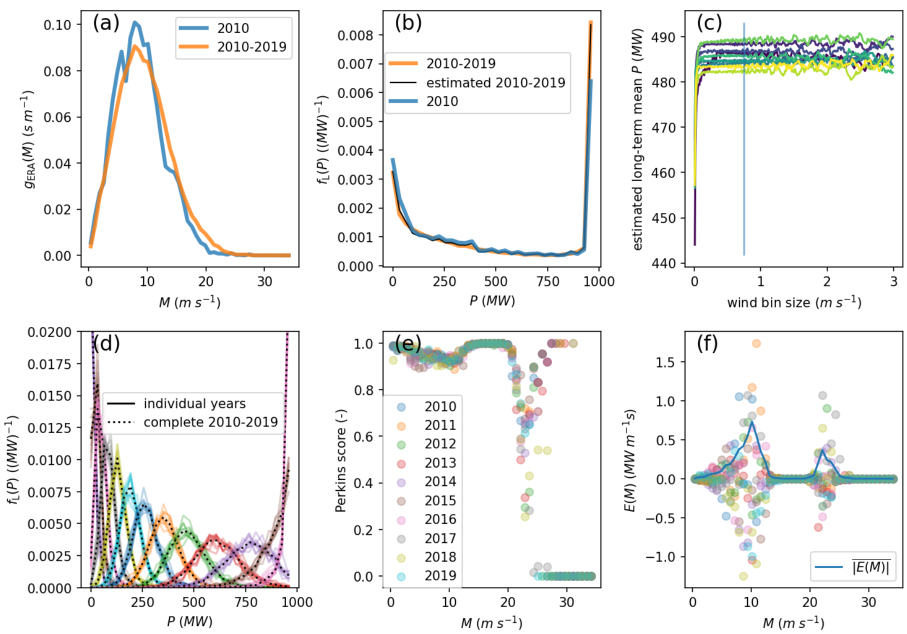

Figure 1: Updated version of Fig. 4, including Perkins Skill Score and $E(M)$.

total final error. For our specific purposes, this makes it a useful metric. Therefore, we intend to add a figure that shows, for each day selection technique and sample size, the absolute value of $E(M)$ (here shown in Fig. 2).

3. Our relatively coarse resolution indeed is a between accuracy and computational cost. Apart from adding a resolution sensitivity study (see our reply on the first community comment), we will emphasize this more in the manuscript.

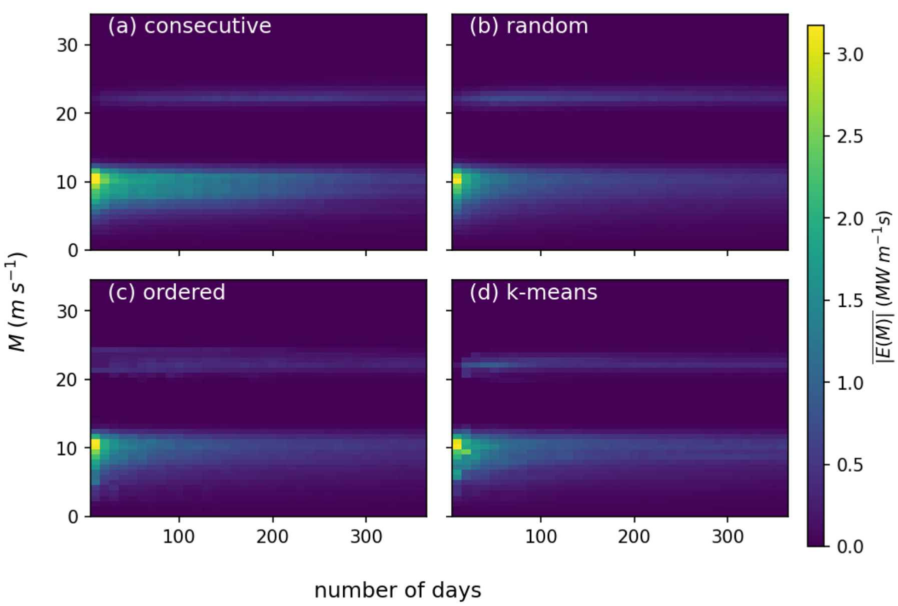

Figure 2: Intended new figure showing the absolute value of $E(M)$ per wind bin and per sample size, averaged over the 500 different samples. This quantifies the source of the error caused by the imperfect match between the short- and long-term conditional probability densities.