Thank you for the additional feedback. It was very well received and we agree that additional details were needed in those two section. Below, please find the edits that we have made beginning a little before the sections mentioned for complete context.

Reviewer comment:
It is great that the authors did add a mention of the IEC standards, saying that their approach is fully adaptable to the standards. I think however that they need to go a bit further and directly discuss how their approach should be adapted to the IEC. The vast majority of such test campaigns are for certifying a certain product or functionality, and then compliance with the standard is a requirement which will take precedence over many other considerations.

Response:
The method described and demonstrated herein is highly flexible and adaptable to the particular needs of the experiment. At a very high level, it consists of performing a suite of simulations to represent a proposed experiment with a balance between computational time and fidelity. The outputs of the simulations are then used to perform a statistical analysis to quantify uncertainty and convergence to standards determined by the user and this data is finally converted into a prediction of the minimum measurement and experiment durations required to produce significant and converged results. At this level, the proposed method could be used for a variety of experiments in many fields, though the focus here is on wind energy and, in particular, field experiments as these present a particular challenge with long measurement durations required to reduce uncertainty due to random errors.

It should also be acknowledged that there are IEC standards relevant to wind energy field experiments \citep[][]{61400-12,61400-13} that researchers may choose to follow. The method laid out herein does not explicitly follow these standards, but it is entirely adaptable to comply with them. If, for example, one wished to follow IEC 64100-12-1 to create a power curve according to standards, then it would be necessary to use the method of bins for uncertainty analysis with the simulated data as detailed in Annex E of that standard. As this is a virtual experiment, however, some assumptions may need to be made regarding the many sources of uncertainty that are tracked and included by the standard but that are not explicitly represented in the virtual experiment. The Category B uncertainties in IEC 61400-12-1 could be to help define an appropriate range of simulation input parameters, for example on wind speed, shear, air density, etc. Uncertainties that cannot be included in estimating input parameters can be included in post-processing of the data. In fact, by including reasonable estimates of every source of uncertainty,, it would be possible to rank the importance of each source through an uncertainty quantification and thereby determine which may be most critical to reduce.

Reviewer comment:
Section 2.2: as recommended, the authors have included statements regarding how simulations do not represent the full variability of the inflow. While this is good to start with, I do not fully agree with the example given by the authors. While TurbSim indeed will drive the distribution of wind towards Gaussian, I believe increasing the simulation period above 10 minutes will lead to relatively small changes in the variability of the flow, because the Veers turbulence model (and other similar models like the Mann model) do not have physical mechanisms that produce turbulence energy with low frequencies. If the tails are longer this is mainly due to more data, the distribution still being Gaussian. This is contrary to physical measurements and more advanced simulation frameworks such as LES. Therefore, more simulations per bin or longer simulations will generate some variability, but not all that is present in the measurements. Another major variability we see in measurements is due to the uncertainty in the inflow characterization, because the wind field is not fully observable (the cup anemometer may have measured 9.8m/s, but the mean speed over the entire rotor may have been 10.5m/s for example). Please discuss this rather than the Gaussianity of the flow which I think has little effect here. How can the lack of wind measurement

uncertainty affect the validity and the usability of the outcomes of a virtual experiment? This is the important question that should be addressed.

Response:

After selecting the simulation method and having acquired representative inflow data, the inflow data are now processed into the format required by the simulation code. Here, the method uses 10-minute bin intervals, which is standard for wind energy field experiments, though it could be easily adapted for other needs. This accepts that the effects of phenomena happening on shorter time scales could be reduced due to long averages and phenomena happening on longer time scales may not be adequately captured, so this averaging time is an important consideration depending on the goals of the experiment. Indeed, numerical representations of inflows will almost certainly underrepresent the true variability in the inflow. TurbSim, for example, will drive the velocity distribution toward a Gaussian and longer simulation times generally create longer tails within the extremes that the model can capture, which will to a point capture a more complete representation of the inflow. If the QoI is an extreme that the model can capture, say, a maximum load, then bins longer than 10 minutes may be necessary such that this QoI is recorded relative to the mean conditions upon binning by condition (binning by condition will be discussed below). If, however, average quantities are of interest, then more 10-minute bins will generally help make up for missing the tails of the distributions of any inflow parameters in each bin.

While more simulations per bin and/or longer simulations will help to replace some of the variability missed when comparing modeled inflows to measurements, it will not close the gap entirely. As mentioned in section \ref{SimMeth}, the proposed method will only yield meaningful results if the modeling tools can capture the QoI, which will require input from subject matter experts. If the QoI is believed to be sensitive to inflow fidelity, then comparisons could be made against higher fidelity methods, such as large-eddy simulation (LES), to verify the adequacy and/or quantify the uncertainty of the low-fidelity approach. These uncertainties can then be incorporated into the final analysis.

Some uncertainties, however, such as the difference between measurements at the met tower and conditions at the rotor are important to retain in the virtual experiment as they can help replicate the real experiment. For example, the velocity measured at the met tower may be biased from the velocity at the rotor. In the control and treatment scenario presented here, this bias is inherently subtracted out. When there is not an available control, such biases in measurements would be critical to capture in the simulations or to incorporate into the post-processing and analyses of the data. Representations of uncertainties in the inflow measurements themselves can and should be included in the uncertainty analysis of the virtual experiment.