Reviewer comment 2, reviewer 1

1. Original reviewer comment: The main area where the paper can be improved is showing how well the wake models predict the change in power or energy from wake steering. The results focus on the ability of the different model variations to predict the power ratios of different sets of turbines relative to unwaked turbines, which is useful for general wake model validation. But since the goal of the models is to optimize yaw offsets for wake steering to increase power capture, the ability of the models to predict the change in power with wake steering compared to normal operation should be investigated in more detail. For example, Figs. 12 and 13 show the power ratios as a function of wind direction for two waked turbines with and without wake steering, but the expected change in power from wake steering is probably small for these turbines and it is hard to determine how accurately the models capture this change from the separate plots. Plots showing the measured and predicted difference between the power ratios for wake steering and normal operation cases would be more effective at showing how well the model captures the change in energy from wake steering. Further, similar power ratio plots for upstream turbines (where we'd expect a power loss from wake steering), combined upstream and downstream turbines, and larger clusters of turbines (e.g., turbines 9-12 and 19-24) could help validate the models in a wider variety of scenarios capturing both power losses and gains at different turbines from wake steering. Aggregate metrics such as the total power or energy gain over a range of wind directions could be helpful too.

Author response: This is a valid point. It is true that the ability of the model to predict wake steering changes is not tested thoroughly in this study, i.e., the ability to predict change is implied from the ability to predict both conditions in the absolute. We have one time series in figure 17 to show that the model can predict power and power ratio well in steering cases, but have not shown its effectiveness at predicting the change due to steering, nor have we shown the ability to predict the sum of upstream and downstream turbines to analyze the overall gain from wake steering. This particular dataset unfortunately does not have enough steering operation to answer that questions rigorously, but is the focus of ongoing work as the system has accumulated more time in operation. Text has been added to the manuscript to explain the need to test this in future work.

It may be true that the particular dataset doesn't include enough wake steering operation to rigorously determine how well the models capture the change in power from wake steering. But some of the existing figures suggest that examples could be shown. In Fig. 11, the power ratios of a waked turbine with and without wake steering are compared to predictions from the initial model (revealing deficiencies in the initial model). A similar plot, comparing the SCADA power ratios with and without wake steering to the best-performing, more sophisticated model(s) would be insightful and illustrate how much better the more advanced models are at predicting the change in power from wake steering. This could be plotted using the data already contained in Figs. 12 and 13. The power ratios in Figs. 12a and 13a could be combined in a single plot to show how well models 1-4 predict the change in power from wake steering (though only the best-performing model(s) could be shown to reduce clutter in the plots).

We have added figure 14 to show these results from all models, for both wake steering enabled and disabled. We believe that this helps illustrate the improvements compared to figure 11.

2. Original reviewer comment: Pg. 1, ln. 18: Another reference to consider citing for the magnitude of predicted wake losses is Lee and Fields (2021): https://doi.org/10.5194/wes-6-311-2021, which shows typical losses between 5% and 20%.

Author response: This reference has been added to the paper.

Reviewer comment 2: Although the introduction says that onshore wake losses for US wind plants have been estimated to be between 2 and 20%, it is worth noting that the values up to 20% in Lee and Fields (2021) are not necessarily for onshore plants.

We have clarified the wording to ensure it is clear that 20% loss is not necessarily for onshore plants, but can include offshore as well.

3. Original reviewer comment: Section 2.4.1: When estimating the Cp and Ct curves using nacelle wind speed measurements, do you account for potential biases (that are also potentially turbine-specific) between the nacelle wind speed measurements and the true freestream wind speed (i.e., by determining and then applying a nacelle transfer function)? Can you discuss how these biases could affect your estimation of the Cp and Ct parameters?

Author response: If biases between the nacelle wind speed measurements and the true freestream wind speed are not turbine-specific, our fitting procedure for CP will automatically account for them. This is because we are using the nacelle anemometer readings themselves to determine both CP and to estimate the ambient wind speed. This was part of our motivation for deriving the CP curve from data rather than using that provided by the manufacturer.

If biases are turbine-specific, this approach could run into issues. However, as noted in the paper, fitting at an individual turbine level was generally impractical due to limited data for fitting at each turbine. Since all turbines were of the same model, it is less likely to pose a significant issue.

We have no independent way to verify the CT curves, and thus we are assuming that the wind speeds in the thrust coefficient tables match those from the nacelle anemometers.

We have added additional content to the paper to briefly discuss these points.

Reviewer comment 2: I agree that any Cp estimation errors caused by biased nacelle wind speed measurements would "cancel out" when calculating turbine power using the estimated Cp and nacelle wind speed measurements. However, it is worth noting that the Cp estimates themselves may be unrealistic. For example, if the nacelle wind speed measurement is much lower than the true freestream wind speed, the estimated Cp could be unrealistically high (> 0.59). Still, as you mention, this would not affect the accuracy of the final absolute power estimate.

Additionally, in Fig. 5, are the legend labels switched? The curve labeled “from manufacturer” is “noisier” than the other curve, but I would expect the historical data-derived curve to be noisier.

A comment has been added to the manuscript regarding the possibility of unrealistic Cp values.

The legend labels are indeed correct, though there is a minor typo in our transcribed power curve at 6.5 m/s. This has been left as-is and noted in the figure caption since this was the power curve input to the initial deployed model.

4. Original reviewer comment: Pg. 17, ln. 392: “but with the target turbine’s wind speed and power measurements omitted”. Are the 1-minute lagged power measurements at the target turbine omitted as well, or are they still used as features?

Author response: The 1-minute lagged power measurements at the target turbine are still used as features. The goal here is to capture any rotor inertia effects that are missed in the steady state approach. However, as you can guess, there is some risk that the model simply uses the previous power measurement as the next prediction! This does become an important model features, but we do see evidence that the model is not simply time shifting the signal.

Reviewer comment 2: Thanks for the explanation. I would simply suggest clarifying that the 1-minute lagged power measurements at the target turbine are still used (i.e., only the current measurements are removed), for example by adding “but with the target turbine’s”current 1-minute rolling average” wind speed and power measurements removed” (or similar).

The manuscript has been updated with this clarification.

5. Original reviewer comment: Fig. 18: Please discuss this plot further. Is model 4 only trained with yaw angle magnitudes > ~8 degrees, so the model reverts to the EWM model for smaller yaw angles? Why is the predicted power gain so much higher for model 4 than model 3? The gain above 1 appears to be an order of magnitude higher with model 4 than model 3.

Author response: There is a lot going on behind the scenes in this plot–thank you for pointing out the lack of clarity. We've updated the discussion to add additional details.

Reviewer comment 2: The added discussion indeed helps clarify the plot, especially how for yaw angles when the output corrector is disabled in Model 4, other turbines which are not affected by turbine 13 still use the output corrector. I had originally thought that the output corrector would have been disabled for all turbines in the farm.

My understanding now is that in the yaw angle sector from ~-6 to +7 degrees, the output corrector is used and outside of this sector Model 4 reverts to the EWM-only predictions for turbines influenced by turbine 13. Since I found this confusing initially I would suggest mentioning in the figure caption which yaw angles correspond to Model 4 using the output corrector and which correspond to the model reverting to the EWM predictions.

This suggestion has been added to the figure caption.