Response to Referee

We greatly appreciate the time taken by the referee to read our manuscript. We have taken into consideration and addressed all comments, questions, and suggestions from the reviewer, and we feel that the revised manuscript is now substantially stronger as a result. Changes made to the text at the request of the reviewer have been highlighted in red in the revised manuscript. In the following, reviewer comments are repeated in italics and our responses are provided in the regular sections of text.

"One final comment is that I personally found the 'leave-one-out' analysis results to be unclear, especially since the error units are dimensional. It is not immediately clear from Figures 9 and 10 whether the GP is indeed reliable (e.g. 0.1 MW error in Region 2 operation when only a single training point is excluded appears to be significant). Perhaps consider reformatting the plot or contextualizing these errors."

We added contextual information to the appendix. We now note that many of the sampled errors are less than 0.1 MW and 0.1 MN-m, particularly in the region of the discovered Pareto set, which correspond to 3% of the maximum power and 6% of the minimum loading, respectively. We also added a note that the multifidelity approach yielded higher maximum errors and lower minimum errors than the single-fidelity approach.