The article is interesting, well written, and deals with the field experimentation of a cooperative wind farm control strategy that aims at increasing the overall power of a wind farm through an appropriate reduction of the power produced by the upstream machines, whose effect is to reduce their axial induction and thus to mitigate the speed deficit in their wakes. Precedent experiments conducted in wind tunnel and simulated environment (properly cited by the authors) have shown that this method seems to be not very effective. As well as the previous experiments conducted in real life have shown extremely small gains in the order of the measurement uncertainty, leading to inconclusive results.

This paper has the merit to describe in very detailed way the experimental setup and the methodology used for the synthesis and the implementation of the controller, as well as the simulation model used to the purpose. For this reason, I think that it is worthy to be published in the journal.

As stated by the authors, the data obtained through the experimentation are not sufficient to derive statistically robust results. Moreover, the used simulation model shows a partial agreement with the experimental data, especially regarding the predictions for the machines located more downstream in the array. The differences between the model-predicted and the experimental overall-cluster power (not shown in the paper, even if they should be) are probably superior to the power gains measured experimentally, a fact that suggests that the model cannot be used to validate the obtained results.

I believe, therefore, that the results obtained do not allow to conclude on the effectiveness or not of the tested method in terms of boosting the wind farm power output. This aspect should be emphasized in the conclusions, which should also include what should be, according to the authors, the actions required in order to reach a satisfactory conclusion on the effectiveness or not of the proposed method.

I also report in the following some other suggestions for improving the manuscript.

- Page 1, line 21-22. The control concepts the sentence refers to are not introduced before.
- Section 3.2 what exactly the control set-point is? Is it the power reduction, expressed in percent of the available one? I think it is important that the authors clarified this aspect. Moreover, I think it is important to show here some resulting LUT, and quickly comment how close the computed optimal set-points are with respect to those adopted by other authors whose findings have been cited in the introduction.
- Page 18-Line 359: to allow the reader understanding why what is shown in Fig 14 is a "small set-point", it must be clarified before what the set-point is.
- Page 26, Line 669-470. I would not claim that 20 degrees of difference in the wind direction is a slight difference. The wake-to-turbines interaction with 225 degree wind direction is totally different from the one with 245 wind direction.
- Figure 20. It would be very beneficial to put aside of Figure 20 a similar figure that shows the delta between measured and simulated “power ratio ON/OFF”. There is the space for it, and it would allow the reader to quickly get how good the predictions of power gains are with respect to the measured data.

- Page 28- Figure22, WT13 data. I would not claim that the agreement is that good as it was for WT 38-37 and, partially, also for WT36. In many instants, indeed, the predicted power is more than double the SCADA data. This probably means that the simulation model is overestimating the wake recovery as we look further downstream in the WTs array.
- Fig. 22: it would be very interesting to show here also the comparison between the measured and simulated overall cluster power. The control set-point are indeed derived with the goal of maximizing the overall cluster power, and LongSim is used as simulation model. it would be therefore very interesting to check how good the model is in capturing this quantity.
- Fig. 22 and 23: I personally don't think that visually comparing time series is the best way to judge the agreement between numerical and experimental data. I would have instead plotted the numerical data w.r.t. the experimental data, also including the correlation factor and the related RMS.
- For what shown in the paper, I would not be confident claiming that LongSim provided an excellent agreement with the experimental data for all WTs. Fig 22, for example, clearly shows that WT13 power is quite often overestimated.