Response to Reviewer #2 - M. Paul van der Laan

The authors develop a meso-scale wind farm parametrization based on the Fitch scheme coupled with an engineering wake model to represent the internal wake losses. In addition, a new wake superposition method is proposed. The predicted internal wake losses are validated with field measurements. Wind farm parametrizations cannot resolve internal wake losses and I think it is very interesting to overcome this by coupling an existing parametrization with an engineering wake model. The article has a focus on the internal wake losses; however, wind farm parametrizations in meso-scale models are mainly employed to estimate wind farm wake losses (for example wind farms situated in a wind farm cluster). Hence, I lack a validation of a wind farm wake produced by your proposed parametrization. You could for example look at the Horns Rev I wind farm wake as used by Volker et al. [1] or simply compare the wind farm wakes of the employed models (new and old) for the present idealized setup. More detailed comments are listed below; they need to be addressed before the article can be considered for publication in Wind Energy Science. Note that I cannot provide detailed comments on the numerical setup of the meso-scale models due to a lack of experience regarding this type of model. Also note that I have not yet looked at any other reviewer's comments before uploading this document because I want to provide an independent review.

The reviewer is absolutely correct that this work focuses on improving WRF and MPAS performance on internal wake losses by incorporating the Jensen wake model. Our study validates the models by comparing the predictions of internal wake losses versus field measurements (observational data) since investigating inter-farm interactions was not within the scope of this work. However, since inter-farm wake interactions originally initiate at the turbines, we believe that, if a model's performance in predicting internal wake effects improves, its inter-farm wake predictions would also improve.

In addition, as the reviewer notes, wind farm parameterizations like the Fitch's one are often used to study inter-farm wakes. But it is important to point out that the inter-farm wake is not treated by the parameterization, but rather resolved directly by the WRF dynamics. As such, the wind-farm wake is going to be treated exactly the same whether the WRF is equipped with the Fitch or the Jensen parameterization.

Main comments

1. Line 26: You mention that mainly two numerical approaches are employed to study wind turbine wakes: LES and WRF. However, there are a lot more numerical approaches to simulate wind turbine interaction, each model has its application area and purpose. Ranking steady-state wake models from low to high model fidelity one could think of: engineering wake models, simplified Reynolds-averaged Navier-Stokes (RANS) models as 2D RANS, parabolic RANS models [Iungo et al. (2018) [2]], linearized RANS (FUGA); and full 3D RANS with actuator disks (AD)[3]. Then we have transient wake models: Dynamic wake meandering model, unsteady RANS with actuator lines, LES-AD, LES with actuator lines (AL), WRFLES-AD and WRF with a wind farm

parametrization. Many of these model are described in G¨o¸cmen et al. (2016) [4], which you already refer to.

We meant to say "CFD," not just LES. All versions of RANS and URANS models, as well as LES, can be categorized as computational fluid dynamics (CFD) models. So, to fix this, we added the references and revised the paragraph as:

"Mainly two numerical approaches are employed: computational fluid dynamics (CFD) and mesoscale modeling. Examples of CFD are Reynolds-averaged Navier-Stokes (RANS) models with various levels of sophistication, from 3D with actuator disks (van der Laan et al. (2015) to parabolic (Iungo et al. 2018), linearized, unsteady, or 2D (see Gocmen et al. (2016) for a review), and large-eddy simulation (LES) with actuator disks or lines."

2. Line 74: Here you mention that the Jensen model works well for any wind farm layout based on the work of Archer et al. (2018) [5]. I do not agree with this strong statement since the performance of the model heavily relies on the wake superposition method and wake expansion coefficient. For example, the Jensen model does not work well for a tightly packed wind farm as Lillgrund and a below-rated inflow wind speed when the original quadratic wake superposition is applied (M2 method in the present work) [6], as you also show in the results section. I could not find the chosen superposition method in the work of Archer et al. (2018) [5], which makes is difficult to understand the results in that work. I think it would be better to write that the Jensen model could be calibrated to get the desired result for a given wind turbine type and wind farm layout.

We used method M2 in Archer et al. (2018) (it was not called M2 in that paper; please see equation 3 of that work on page 1189). In this paragraph, we mention two reasons for choosing Jensen: 1) it is possibly the most widely used model, and 2) it performs reasonably well regardless of the wind turbine layout or wind direction. We do not mean that Jensen is always the best model. Still, compared to other models, it appears more successful overall when applied to various wind farms using a wake expansion coefficient of 0.075 and 0.04 for onshore and offshore wind farms, respectively (Archer et al., 2018). According to our previous investigations, every wake loss model's performance strongly depends on the lateral and axial spacing between turbines. The more packed the wind farm is, the less accurate the models are. All models we have studied (and not just Jensen) show the worst performance in terms of absolute error and bias at packed farms (such as Lillgrund) and the best performance at the most widely spaced farms (such as Anholt). According to the extensive study by Archer et al. (2018), the Jensen model appeared to be one of the two models that showed a consistently strong performance among all types of wind farms (packed, moderately-spaced, and widely-spaced ones) and for all directions. Again, although in some cases it was not the best, it never appeared to be the worst either, and overall, it appeared to be one of the safest models one could choose.

With that said, we moderated the paragraph as:

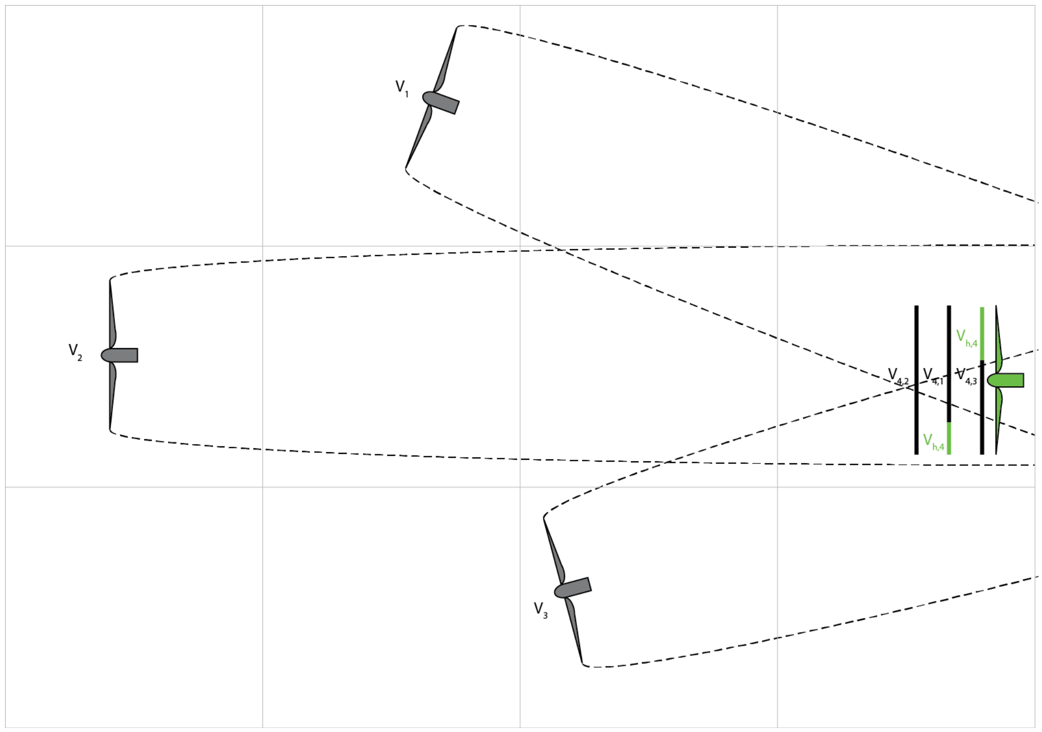
"The Jensen model was selected for this parameterization … because it performs reasonably well regardless of the wind turbine layout or wind direction. … While the Jensen model was not the best model all the time, it stood out for its consistently strong performance and for rarely ranking last for all directions and all farms."

3. Section 2.2.2: It is interesting that you investigate different superposition methods. What is the physical meaning of method M4? I find it strange to superpose the absolute wind speed instead of the wake deficit. It means that you take a L2 norm of the wake wind speed normalized by the number of upstream turbine wakes, which is a form of averaging rather than a summation or superposition method. In other words, method M4 will always smooth out the wake effects between the largest and the smallest (upstream) wake wind speeds. There are also situations where the superposition method does not make sense. For example if one has a regular layout with a large lateral spacing (in y) but a small stream-wise spacing (in x), such that the turbines do not interact laterally, then a downstream wind turbine, will still experience reduced wake effects from the lateral turbines because you divide by the number of upstream turbines N.

In the example of large lateral spacing and small streamwise spacing, M4 will not behave as described by the reviewer. The Jensen model predicts a finite wake laterally. If the lateral spacing is large, then the wake will not expand laterally enough to hit the turbines downstream, thus there will be no wind speed reduction (N=0, thus no wind speed reduction is applied). Basically, only the turbines that have a wide enough wake to inpact turbine i will be accounted for in the value of N. N is the number of turbines whose wakes impact turbine i, not just the number of upstream turbines or the number of turbines in a grid cell. This was stated at l. 163: "When multiple wakes from multiple turbines j (j=1...N) overlap at turbine i".

Having said this, the reviewer is correct that M4 will always give you basically an average of the wind speeds of all the wakes involved, which by definition will be higher than the lowest wind speed (and lower than the highest). This makes the most sense when the turbines are actually not aligned with the wind direction, which is the case for which this method was designed (multi-cell).

To understand this, let's look at the figure below, in which each turbine is in its own grid cell, and the wind directions are slightly different in each grid cell, as is the case in real simulations (also, the wind speeds are different, of course).



Turbine 4 is affected by 3 wakes, each of which causes its own wind speed (e.g., V4,1 is caused by turbine 1 and is a partial wake case, with some of the rotor affected by the wake and the rest unaffected, thus experiencing Vh,4). Suppose V4,1 = 6 m/s, V4,2 = 7 m/s, and V4,3 = 5 m/s. Note that these are the resulting wind speeds at turbine 4; they are not the undisturbed wind speeds at the upstream turbines. The resulting wind speed at turbine 4 will be 6.1 m/s, as a result of the mixing of the three individual wakes.

Why should the resulting wind speed be lower than 5 m/s? If that was the case, neither mass nor momentum nor energy would be conserved.

If the turbines are aligned, then this is less intuitive and may even appear to be incorrect. We are used to seeing the squared deficits from the Jensen's top hat shape overlapping and reducing the wind speed more and more with more overlapping wakes via the M2 method. This is actually not happening in the real world. We know that the wind power and wind speed plateau very quickly at the second or third turbine in a row due to partial wake recovery and entrainment compensating for the momentum extraction. If you look at the figures in our manuscript for directions 222° or 180° at Lillgrund or 339° and 183° at Anholt, you will see examples of observational data from the farms indicating that the third turbine generates more power than the second. With method M2, this trend cannot be reproduced; however, M4 appears to be successful in modeling such cases.

Method M2 (and M3) does not conserve energy or mass or momentum. Why should the squared deficits be conserved? They are a proxy for kinetic energy, but the mass is *missing, and the air is incompressible; thus, density cannot change, and it is impossible that kinetic energy sums up that way.*

In addition, if we assume n equally distant aligned turbines, then the equation for M2 is basically a geometric series like this:

$$\sum x^{2n}$$

with |x|<1. Thus, as n increases, it converges to:

$$\lim_{n\to\infty} \sum x^{2n} = \frac{1}{1-x} > 1$$

which is greater than 1. This means that each additional wake deficit incrementally reduces wind speed and can even cause a zero wind speed or even negative. This is not physical and would never occur in reality. Method M4, however, does not suffer from this problem.

M4 does not conserve mass or momentum or energy either, but it will never result in a negative or zero wind speed and can be thought of as a way to account for entrainment and partial wake recovery. With M4, the resulting wind speed will be the average of the (partially recovered and therefore relatively high) wind speed from the farthest turbine and that (fresh and relatively low) of the nearest turbine, plus all the others in between. M4 effectively avoids the unrealistic continuous drop in wind speed as more and more turbines are aligned that other wake superposition methods suffer from. The fact that it outperforms other methods is also an indication that it is not incorrect.

We added the following additional sentence on p. 7:
"In a sense, M4 can be thought of as a way to indirectly account for partial wake recovery due to the added turbulence from multiple wakes and entrainment."

4. How is the freestream wind speed defined/ calculated in your model? If it is the cell average, then this would mean that you would need to iterate over the Jensen model, since the cell average influences the internal wake model results, which influence the cell average. In case you use the freestream from the inflow, which is possible for an idealized case, how would you extend this to a realistic meso-scale setup where the inflow is transient?

We use the cell average wind speed at the beginning of the time step as the free-stream wind speed. We had not thought of using an iterative process, but we suspect that it would further decrease the wind speed within the grid cell and may or may not converge to a constant. Hence, we apply the Jensen model only once.

We would like to bring up three important items to support our choice.

First, in the current Fitch parameterization, wind speed at every turbine within a grid cell (Ui) is always equal to the grid-cell wind speed (Uh), i.e., Ui=Uh all the time. This means that the current parameterization completely ignores the internal wake effects within the same grid cell. By contrast, the proposed Jensen-based parameterization uses the hub-height grid-cell wind speed (Uh) as the free-stream wind speed in the Jensen model to compute the wind speed that each turbine of that cell would experience.

Second, we acknowledge that Uh does not exactly follow the definition of free-stream wind speed. We struggled with it ourselves, but eventually we concluded that any alternative would be even more arbitrary and would potentially introduce more errors than using the grid-cell wind speed. For example, we could try to identify an upstream grid cell that could be considered a free-stream. But this grid cell would be wind direction-dependent and grid-size dependent. If you have a wind farm that is distributed among dozens of grid cells, each with its own wind direction and wind speed, which wind direction do you even pick to identify an upstream grid cell? If the wind farm is, say, 50 km long, does it really make sense to pick an upstream wind speed (and direction) that is 60 km away? What about the possible induction zone? We concluded that any other choice would be more arbitrary than just choosing Uh.

And third, a parameterization should be easy to implement and fast. It would be impractical to propose a parameterization that relies, for example, on idealized simulations without the farm, or on picking a grid cell outside of the wind farm somewhere. Choosing Uh is straightforward and easy to implement.

We changed the notation at line 160 to be consistent and replaced U(x)/U_inf with Ui/U_inf and added the above discussion to the manuscript as follows at p. 8:

"Alternative choices could be made for Uinf,i, such as the wind speed at a grid cell located at some distance upstream of the wind farm along one wind direction (among the varying wind directions simulated inside the farm), or the wind speed from offline simulations without the wind farm. However, any alternative would be even more arbitrary and would potentially introduce more errors than using simply the local grid-cell Uh. Another advantage is that Uh is straightforward to implement in the codes and fast to execute."

5. How do you calculate the cell average wind speed from the Jensen model? Do you use the cell area only (and thus you would disregard the part of the wakes that extend beyond the cell area) or do you use an area that covers the entire wake of all the turbines within a cell?

A cell's average wind speed, which is used as the free stream velocity for turbines within the cell, comes from WRF (Uh) and is not calculated via Jensen's model. Suppose Uh is the cell-average wind speed at a cell that contains turbine i. This average wind speed calculated at the previous time step is plugged into the Jensen model (Equation 6) as the

freestream wind speed to calculate U_ij, which is the velocity experienced by turbine i caused by an upstream turbine j, which is not necessarily in the same cell as turbine i. After wind speed or wind speed deficits caused by all turbines upstream turbine i are calculated, they are superimposed using a superposition method (methods M1 to M4). That yields total wind speed at turbine i and allows for calculating power production as well as updating u and v (using the exact same equations as those used by Fitch, Eq 1 and 2, with the only difference being a localized wind speed rather than the same one for all the turbines in the grid cell, Eq. 13-14), which would then be used for computing a new Uh at the next time step.
So, the average wind speed within a cell always comes from WRF, we did not calculate it (except for the vertical interpolation, like in Fitch's, to get the value at hub height).

6. Equation (14): Should the left hand side be ∂vk/∂t?

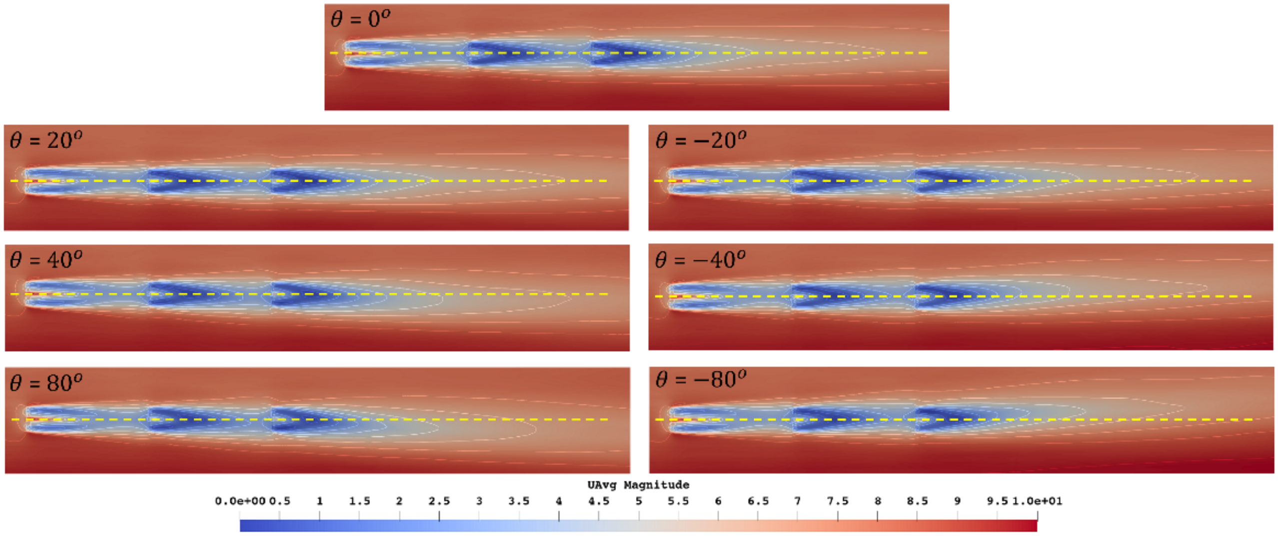
Excellent catch. Thank you! We fixed it.

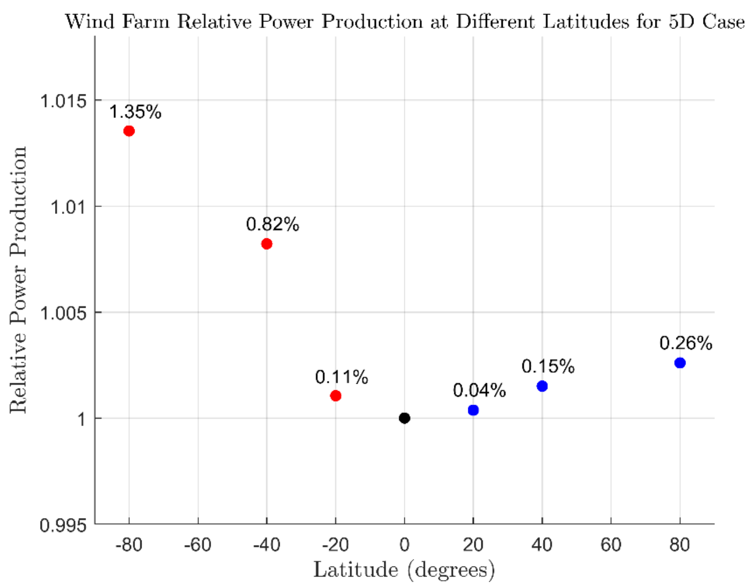
7. Line 240: You propose to use a Gaussian filter with a standard deviation of 2◦, where did you base this value on?

To our knowledge, there is no recommended value of the standard deviation to use in mesoscale simulations and we did not have time to verify the sensitivity of the results to the choice of the standard deviation for the Gaussian averaging. The value of 2° is loosely based on Gaumond et al. (2014), where they showed measurements with a standard deviation of 2.67° (their Figure 2). Since mesoscale models like WRF and MPAS are less variable than the observations and have already a smoothing effect, we wanted a value smaller than the standard deviation of observations and we rounded down to 2°. We added this to the limitations of our study on p. 34:
"In addition, we did not test the sensitivity of our results to the value of the standard deviation used for Gaussian averaging."

8. Line 284: Why do you use a latitude of 50◦ ? Lillgrund and Anholt are located at 55.5◦ and 56.6◦, respectively.

To be honest, we do not recall the particular reason. We probably mistakenly approximated the latitude of the two wind farms as about 50N. We do not think this mistake will have any significant impact on the results. For a previous study, we conducted LES using SOWFA to check the sensitivity of the simulated wakes to latitude and used the entire range from 0° to +/-90°. We computed the variation of the total power production of three inline turbines with an axial distance of 5D (see the following figures). We found that a difference of 5° around 50° has minimal impact on the wake deflection. Using linear interpolation, the total power production changes by only 0.014%.





9. Section 3.1: Why are the measurements not filtered for neutral conditions? Without such a filter it is not fair compare the numerical results for neutral conditions with the measurements. If you prefer to keep the current measurements results you need to clearly state (in results section and conclusion) that the model validation is not entirely fair.

The reason we did not filter the data based on stability in this paper, even though we are painfully aware of its importance and have published several papers about it, is that the datasets we used did not include any information on stability. As such, we could not conduct any stability-dependent validation anyway, so we decided to stick with neutral stability for the simulations. We forgot to mention it and added it on p. 10:

"No atmospheric stability information was available from the measurements at either wind farm."

and:

"Although we recognize that atmospheric stability may impact the evolution, shape, length, and duration of the wakes (Ghaisas et al., 2017; Xie and Archer, 2017), neither the Lillgrund nor the Anholt dataset included any measurements of atmospheric stability and therefore in this study we performed all the simulations under neutral conditions."

and in the Conclusions:

"Due to the lack of atmospheric stability information from the measurements, all the simulations were conducted under idealized neutral conditions."

10. Section 3.2: It seems that you are modeling an idealized setup in the mesocale models. How do the steady-state inflow profiles of wind speed, wind direction and TKE/TI look like for both WRF and MPAS? You could plot these results in Fig. 11. What is the actual roughness length applied by the Charnock relation? What is the turbulence intensity at the investigated hub heights?

The Charnock relationship gives surface roughness from the surface friction u, thus it is a dynamic calculation. There is not one value of it, especially after the turbines are added. To access the values of z_0, we would need to rerun the simulations and select z_0 as an output variable.*

We agree that adding the undisturbed profiles of TKE and/or TI would be somewhat interesting. However, modifying the figures to add this information would be a very difficult task for us at this point. The first author, who was a postdoc working with us at the time of this research and who did all the figures, has moved on to a new life in China and we are having a hard time accessing and understanding his folders and files. Aside from the practical difficulty, we also would like to point out that we do not have any TKE or TI information from the measurements, thus we would not be able to gauge if the simulated profiles are realistic. This means that the additional profiles would not be terribly useful because they could not be verified. Also, we obtain results that are reassuringly similar to those from the Fitch parameterization, which suggests that our results are not incorrect or unrealistic. We could modify this figure by spending more time and resources on it if you find it absolutely necessary, but it would be a daunting task.

11. Figures 3-8, measurements: What is the wind direction bin size used for the observations? In addition, you could consider to plot the error bars as the uncertainty of the mean (standard deviation dived by the square root of the number of samples).

The wind direction bin size was 0.5 degrees. We added this information to the caption of Figure 3.

Regarding the second part of your comment, while dividing the standard deviation by the square root of the sample size may be a better error bar, unfortunately making this small change to the figures would be extremely difficult for us at this point. As mentioned earlier at #10, the first author, who was a postdoc working with us at the time of this research and who did all the figures, has moved on to a new life in China and we are having a hard time accessing and understanding his folders and files.

12. I do not think it is fair to normalize the wind turbine power of Fitch with the power of the first row wind turbine, as Fitch is meant to model the average power of a number of wind turbines (or entire wind farm for the single cell case). You could overcome this by normalizing the wind turbine power of all the models and measurements with a hypothetical free-standing wind turbine power (using the freestream value at the wind farm location [without any wind turbines present] and the power curve. The measured freestream could be evaluated by using the power of a group of front row turbines and the power curve. A similar approach was performed in Hansen et al. (2015) [7].

The definition of relative power is the ratio of the power of each turbine divided by that of the front-row turbine of each row/column. We have used this definition in this and all our other publications. It is not what the reviewer describes. Our treatment is fair because it is the same for Jensen and Fitch. Also, it actually helps visualize that, with the Fitch parameterization, all the turbines in the same grid cell are treated equally as front-row turbines, thus they truly should have a relative power of 1. When the wind farm covers multiple grid cells, the relative power from the Fitch parameterization is no longer 1.

13. Line 478: You mention that the Jensen parametrization is insensitive to grid resolution: The Jensen parameterization tends to underestimate the power, regardless of the alignment or non-alignment conditions and regardless of the grid resolution. The consistent sign of the bias (negative) in column power output and the absence of sensitivity to the grid resolution or wind direction are all desirable properties. However, in order to show a grid independence you need to compare results of at least three different grid levels and show that the results converge with horizontal grid refinement. For example, you can add another grid level in Figs. 9 and 10, for example for 0.5 km or for 4 km. (The 4 km cell size might not be interesting for Lillgrund as this is the same at the single cell approach. (I am aware that grid-independence is challenging for WRF due to parametrizations that rely on large horizontal cells, but it might be possible for an idealized setup.)

We have already run both models, WRF and MPAS, at three resolutions at each farm (24 km at Anholt and 4 km at Lillgrund for the single-cell cases, and 2 km and 1 km at both for the multi-cell cases). The 4-km resolution results are those called "single-cell" for

Lillgrund; the single-cell resolution for Anholt is 24 km because the Anholt farm is too big to be contained in a 4-km grid cell.
Also, we would like to point out that we have run two models for multiple wind directions (6 at Lillgrund and 8 at Anholt) and multiple wake overlapping methods, for a total of over 100 simulations.
We agree though that the term "absence of sensitivity" might be too strong. We modified the sentence as follows:

"The consistent sign of the bias (negative) in column power output and the minimal sensitivity to grid resolution or wind direction are all desirable properties."

14. The cases considered in this article only look at high thrust coefficients, since a below rated inflow wind speed is used. The conclusion on the best superposition method might change if you had considered an above rated wind speed, where the thrust coefficients are smaller. This it because the performance of a super position method can be shown to be dependent on the thrust coefficient, see for example Machefaux et al. (2015) [8]. You could add an above-rated case or you could add a comment/discussion in the article.

We added this discussion in the conclusion section as:

"A limitation of our study is that we only considered wind speeds that are below the rated wind speed, thus with a high thrust coefficient; when the wind speed is above the rated wind speed, the thrust coefficient decreases. Since the performance of a superposition method depends on the thrust coefficient (Machefaux et al. 2015), our results might be different in such cases."

15. Line 370: You write: Both M3 and M4 reproduce well the feature of power output becoming steady after the fourth turbine in an alignment column (Fig. 4). I understand what you mean, but I think it is better to write about a balance instead of a steady-state. Here, I mean a balance between the momentum extracted by the wind turbines and the momentum transport into the wind farm from the boundary layer.

We agree. The statement was revised as:

"Both M3 and M4 reproduce well the feature of power output remaining unchanged after the fourth turbine in an alignment column (Fig. 4), due to the balance between the momentum extracted by the turbines and that replenished from the boundary layer."

16. Figure 11: I would normalize the wind speed by the freestream wind speed, U∞, and the TKE by p 2/3TKE/U∞. The height can be normalized as (z − zH)/D, with zH and D as the hub height and rotor diameter, respectively. The same applies to Figs. 12 and 13, where the cell power production could be normalized by the total wind farm power.

We agree that presenting data in a dimensionless format is beneficial when studying fundamental wake or turbulence behaviors. However, for practical applications like this one, we think that showing the actual wind speeds, the actual wind speed deficits, and the actual TKE is more valuable than the normalized versions.

17. Line 496: You mention that a wind direction of 222◦ results in the largest wake losses for Lillgrund but this should 120 or 300◦ , where the wind turbine spacing is smallest.

The paragraph was revised as:
"For these wind directions, the wake effects are neither strongest nor weakest, which allows for a representative comparison."

18. Line 547: When referring to the Gaussian wake model, I think you need to refer to the original work of Bastankhah and Port´e-Agel (2014) [9].

Done.

Minor comments

1. Thanks for referring to my work. My last name is written with small letters for the first two words: van der Laan.

Thank you for correcting us. We fixed this in the revised manuscript.