We first like to thank the reviewer for the comprehensive, fair and constructive comments. Please find our answers to the comments below. We marked the reviewers comments blue and our answers black.

It should be noted the very large computational demand to achieve these results: each simulation comprises 2088 wind turbines with mesh sizes ranging from ∼ 7.4 × 109 to ∼ 1 × 1010 cells, which is indeed remarkable (so it will be very informative to reveal the computational expense).

In line 134 we added a sentence on that:
"The simulations were carried out on 5120 cores on one of the supercomputers of the North-German Supercomputing Alliance (HLRN). A simulation required a Wallclock time of 25 to 50 h."

I recommend this work for publication but please
provide an answer to the following questions and comments:

L89 “it is possible to prescribe a surface heating . . . ” is this just a choice or a necessity, in view of the issues with the direct prescription of the surface flux as described in Basu et al. (2008)?

The importance of prescribing a surface heating rate instead of a heat flux for SBLs (and CBLs) is highlighted later in sections 2.4.2 and 2.4.3.
In section 2.4.3 (SBL) we rephrased line 241 to:
"A Dirichlet-condition is applied for the surface temperature, because prescribing a surface heat flux can lead to unphysical results (Basu et al., 2008)"

L94, Could you please briefly comment why the overestimation and how this is fixed? Also, would the use of a different distribution length along the smearing direction help to remedy this (a different value of the standard deviation than 1Δx in the Gaussian distribution)? I think it is useful to provide these details here (even crucial in the case of the overestimation) as the reference you provide is not in English. Also note that in the archive of Meteorologische Fortbildung at the following link, the issue of the reference you cite appears as from 2015 instead of 2016: https://www.dwd.de/EN/ourservices/pbfb_verlag_promet/archiv/archiv_promet.html

Yes, we kept this part short. We have now added some information and changed the year to 2015:

"The ADM is described in detail by Steinfeld et al.(2015) and Wu and Porté-Agel (2011). To avoid numerical instabilities, the disc element forces are distributed to the neighboring grid points by a three-dimensional Gaussian smearing kernel, which is approximated by computationally less expensive 4th-order polynomial. The smearing kernel has a default radius of 2Δx, reaching approximately 78 grid points. The otherwise two-dimensional actuator disc is enlarged in the axial and radial direction by the smearing, resulting in a power overestimation of 26.8 %. The power overestimation can be reduced to 12.5 % by setting the kernel radius to 1Δx, reaching approximately 10 grid points, without any numerical instabilities. The thrust coefficient is overestimated by 2% for 2Δx and underestimated by 4% for 1Δx. As a compromise, the smearing kernel radius is set to 1Δx for this study. The wind turbine power output is corrected for the power overestimation by a factor of 1/1.125 before entering the wind farm power output analysis."

L127 and Table 1: should one assume that 225 deg is the wind direction at hub height obtained with a ug at direction alpha as shown in the table?

Yes, it is hub height wind direction. We added that:

"The wind direction at hub height is set to 225° by tuning the geostrophic wind direction \alpha appropriately (cf. Table 1). Southwest-wind is one of the most common wind directions in the German Bight."

L133, why have you chosen to use a uniform cell distribution in the horizontal plane? This question arises from the fact that if cell stretching was used away from the turbines, the savings could have been used to achieve a higher refinement either around the rotor and wakes or for the near ground zone, where the mesh requirements for a well-resolved LES are higher.

PALM allows only for a vertical stretching. In the horizontal directions the cells have to be uniform. The nesting feauture of PALM could have been used for sparing resources, but we decided against this option in order to avoid any numerical effects on our results that might be introduced by the nesting. Thus we decided to use this uniform grid with 20 m grid spacing.

L134, why is this resolution enough? In the absence of a numerical or experimental evaluation, please provide an argument. This element is crucial as the recovery of the individual wakes depends on the amount of TKE resolved within and around.

There is a numerical evaluation in Steinfeld (2015) that shows that 5D behind the rotor (which is the smallest turbine spacing used in our study and due to the staggered configuration the first wake-turbine interaction happens at 10 D) the wind speed profiles for a grid spacing of below 10 grid points per rotor diameter are the same as for finer grid spacings. They used the 5MW NREL turbine with D=126m and results with 16 m grid spacing were still good at 5D, Fig. 4-7.

We added this information:

"This grid spacing yields a density of 12 grid points per rotor diameter, which is enough to resolve the most relevant eddies inside the wind turbine wakes. As Steinfeld (2015) showed, even 8 grid points per rotor diameter are sufficient to obtain a converged result for the mean wind speed profiles 5D behind the turbine."

L134 (and L204), also with regard to the cell resolution, please comment on its the adequacy in terms of the amount of subgrid TKE with respect to the total. In other words, how "well resolved" is the LES, especially for the wake region? L204, please describe how the subgrid convergence study demonstrated that 20 m was the adequate cell resolution in the horizontal directions.

The ratio of SGS-TKE to total TKE and SGS-momentum flux to total momentum flux is smaller than 10% for the SBL precursor run.

We added this information:

"Test simulations with 10 m and 20 m grid spacing showed that a grid spacing of 20 m is sufficient, if this SGS-model is used (less than 1 % difference in wind speed maximum and less than 5 % difference in BL height), whereas the results are more grid spacing sensitive (2 % difference in wind speed maximum and 20 % difference in BL height) if the standard-SGS model of PALM is used. For the SBL precursor run the ratio of SGS-TKE to total TKE and SGS-momentum flux to total momentum flux is smaller than 10%, except for the lowest grid point."

For SGS-TKE in the wake see the comment on line 386.

L165 and results section: it is mentioned that blockage from the parks is observed to reach the recycling plane, did the authors observe any dependence of the blockage with respect to the size of the domain (regardless of the fact that the boundary conditions at the sides are periodic)?

We have not tested this for this setup. But we have carried out tests for different domain length (Lx) and width (Ly, periodic) for smaller wind farms and domains. The effect on the wake flow and the flow blockage was negligible.

L168, it's not totally clear how the height limitation set to the recycling of turbulence helps to avoid the increase in the BL height, is the subsidence mentioned in L90 not sufficient? However, it is clear that this region represents only a small portion of the total domain length, albeit still large (10 km).

We think that this sentence is clear. If no turbulence is recycled to the inflow above a certain height, then the flow is laminar above that height and thus the (turbulent) BL is limited to that height. In the CBL- case the BL growth is compensated by the subsidence. However, the BL could potentially slightly grow before the recycling plane due to the blockage effect of the wind farms. This growth would be mapped to the inflow and reinforce itself without the above mentioned recycling limit.

L194, please clarify the usage of the concept "steady state". It seems that in some instances the authors use these words when referring to a flow that has achieved statistical convergence such as in a developed BL which could be confusing with a flow that is indeed steady-state and as such, lacks transient features such as those reproduced by LES.

Also due to a comment of the other reviewer we rephrased and added information:

"While steady-state turbulence is reached after only a few hours, achieving a steady-state mean flow can take several days. Here, we declare the mean flow as steady if the oscillation amplitude of the hub height mean wind speed is less than 0.5 % and declare the turbulence as steady if the change in friction velocity is less than 2 % in 4 h."

L230,235, could the issues pointed at by Basu et al. (2008), such as those related to the estimation of the friction velocity, be added to these arguments?

If we understood it correctly, then the issues pointed out by Basu et al. (2008) are only valid for SBLs. This section (L230 ff) deals with a CBL.

L378, is it possible to provide an example (or a threshold) of a park size where, following these arguments, the turning would occur in the clockwise direction?

We think that deriving such a rule is not possible without performing additional simulations. It might also depend on stability and BL height, because the clockwise turning is caused by wind veer. Respective studies will be addressed in a follow up study.

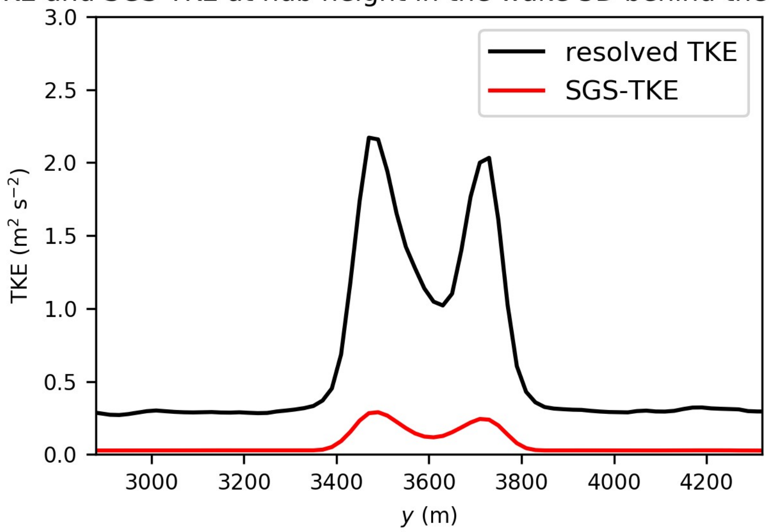
In Figs 5 and 6, scale is inverse with respect to the other (red to green and green to red), please maintain consistency to denote low to large values in scaling. Furthermore, it is rather common to see blue to red to depict small to large values (e.g. subfigure (f) in figures 4, 6 and 7), so to use blue in the middle can be somewhat disconcerting. Is this perhaps a choice to highlight the wakes behind the clusters?

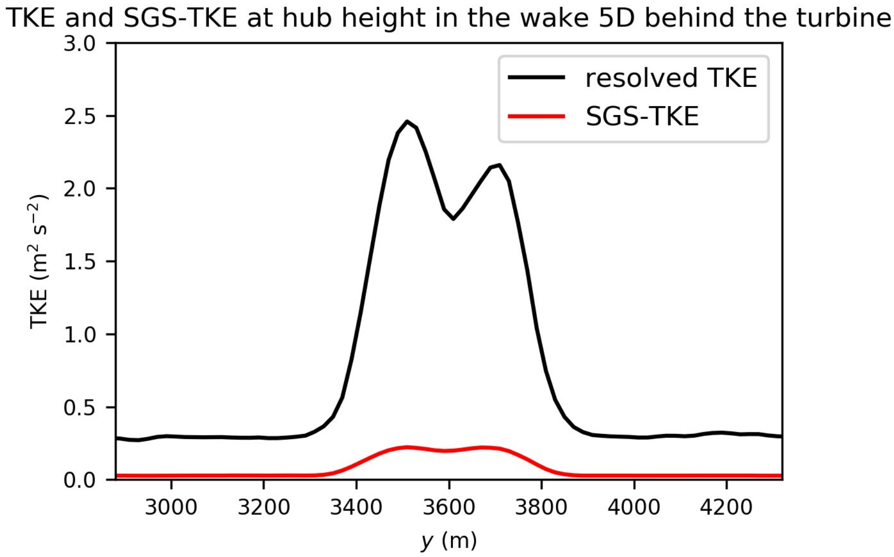
This choice is intentional. Blue and especially red should indicate the wake, which is smaller wind speed but higher TI. We put blue in the middle to avoid a green-red color transition which might be difficult to see for red-green colorblind people.

L386, the TKE does consider the subgrid scales. Since the resolution is relatively coarse to re-solve the flow around the rotors, a sizable part of the TKE is potentially found within the subgrid scales. Please provide an argument regarding why only resolved scales should be considered in this analysis.

Resolved and SGS-TKE is displayed in the following figures and justifies that the SGS-TKE has no significant contribution here.

TKE and SGS-TKE at hub height in the wake 3D behind the turbine





In line 387 (equation for TI and TKE) we added:

"The SGS-TKE is neglected, because it is smaller than 10 % of the resolved TKE, inside the wind turbine wakes at a distance of 3D or more."

Figure 7, label in vertical scale should be z.

We corrected that.

L505, how is the power computed for each turbine?

The power is part of the output quantities of the advanced actuator disc model. We added a reference to an article of Wu and Porté-Agel (2010), which describes details of this model (in English language) in line 94 so that the interested reader can refer to this article.

DOI: 10.1007/s10546-010-9569-x

L608, it is remarked that for the SBL-300-7D case, the W gpg,wt component is dominant but in the

Fig. 10(e) it seems that W vke f is about the same value throughout except only within the two gaps between farms.

Yes that is true. We rephrased that to:

"For the first 10 km of the wind farms the vertical kinetic energy flux dominates but further downstream, the energy input by the pressure gradient below rotor top level is greater or equal to the vertical kinetic energy flux"

Technical corrections

• L121, typo, replace "can not" for "cannot"

- L162: I suggest to remove first comma
- L308, typo, it should be “effect” (it says affect)
- L637, typo, “of” instead of “if”

We thank for these hints, and have considered them.