This paper studies the wake properties and power output of a potential future wind farm scenario in the German Bight by means of massive large-eddy simulations (LES). The simulations cover an area of about 200 by 160 km with a 20 m grid resolution and include several wind farms amounting to a total of up to 2088 wind turbines and representing a total wind farm capacity of up to 31 GW, thereby exceeding any other available LES study by at least an order of magnitude. Five different LES cases are meticulously performed so as to investigate the impact of turbine spacing, boundary-layer height, and atmospheric stability. I believe the LES are well-designed and the resulting LES dataset is quite impressive and potentially very useful for the wind energy community. Further, the authors have presented detailed analyses of the dataset and have found several interesting conclusions. The paper is well written and the figures are very informative. Overall, I believe the research is of good quality and potentially very interesting for the community. However, I do have some major concerns about some of the discussions in the paper. I listed my main concerns below together with other scientific questions and minor/technical comments.

Main concerns:

1. The authors have a tendency to state physical explanations of observed flow behaviour as a fact rather than as a hypothesis or backed-up by specific analysis. While the explanation is often plausible, the authors should more clearly indicate when they are presenting a hypothesis or whether they actually have evidence supporting their claim. For example
    - Line 166: "Due to the self-reinforcing behaviour of this process, the crosswire variations can build up quickly without y-shift, even if the wind farms have a distance of 15 km to the recycling plane." It is unclear to me whether you actually saw this in preliminary calculations or if this is a hypothesis?
    - Line 295: "This BL height dependency occurs because a thicker BL contains more kinetic energy that can be transported down to the wind turbine level by turbulent vertical mixing than a shallow BL." Do you have evidence supporting this hypothesis? If this is something that will be discussed later on, please mention so explicitly.
    - Line 296: "The wake length and speed deficit of small wind farms (e.g. N-1 and N-2) is relatively unaffected by the BL height because the wind farm induced internal boundary layer does not reach the top of the BL." Did you investigate the internal boundary layer development? If not, how can you know that this is the reason? It would be interesting to add IBL development to figure 7 (and extend the analysis to small wind farms as well, see below) to support this hypothesis.
    - Line 314: "The case SBL-300-7D covers several flow features that cannot be seen in the other cases." Are these flow features not visible because they do not occur or because they are smaller? Did you verify quantitatively whether there is any flow deceleration in front of the wind farm in neutral or unstable conditions?
    - Line 418: "Because the mean wind speed inside the BL decreases in the streamwise direction, the IL must be displaced upwards in order to maintain a constant mass flux inside the BL." Note that the mass flux can also be conserved by means of an acceleration above the wind farm (below the IL) or by airflow to the sides. In your results, the IL is displaced upward, but that does not necessarily mean that this will always be the case (e.g., a stronger capping inversion might lead to flow acceleration or flow diversion to the sides). Please rephrase.

- Line 532: “The wind farm efficiencies for the NBL-case are greater because the inflow wind speed in the bulk of the BL is higher for the NBL-case than for the CBL-case.” Higher bulk wind speed causing higher efficiencies seems a very plausible explanation, but I’m not sure you can deduce this conclusion from your results with 100% certainty. Maybe rephrase?
- Line 596: “As stated earlier, the power output of infinitely large wind farms is determined by the energy input of the pressure gradient. Hence, the power output of infinitely large wind farms does not depend on the BL height, at least for this idealized setups with a stationary CBL inflow. ” Did you actually run simulations of infinitely large wind farms with stationary CBL inflow to confirm this hypothesis?

2. The authors made an effort to cover an enormous area in their simulations, but the bulk of the analysis is based on the large wind farm cluster in zone 3 (e.g., many analyses look at cross-sections at y-120 km). I think it would be useful to include more analyses of the smaller wind farms to be able to contrast the behaviour of “small” and “very large” wind farms, as the difference in flow behaviour between small and very large wind farms appears to be one of the main messages of the paper. For example, figures 5, 7, 8 and 10 focus solely on the very large wind farm cluster and there is no counterpart analysis for one of the small wind farms. Moreover, it is interesting to see in fig 10 that the vertical kinetic energy flux continues to decrease over the wind farm and does not reach a plateau, unlike what was found by Allaerts and Meyers (2017, J. Fluid Mech.). I wonder whether this is because you are looking at much longer wind farms, and therefore a similar analysis for a smaller wind farm would be useful. As a matter of fact, line 615 states that “These results show that the power output and the wake of very large wind farms behave very differently compared to small wind farms.” At least in terms of the power output, you haven’t shown this in the paper even though the data is available.
3. The authors decided to take a simplified approach to the energy budget analysis, but I believe too many terms are left out to support the discussion. A more detailed or even full energy budget analysis seems warranted.
    a. Line 539: “the extraction of kinetic energy by the wind turbines can be compensated for by two sources of energy: (1) Vertical turbulent flux … and (2) Work done by the geostrophic pressure gradient.” This is not entirely correct, as the extraction of kinetic energy can also be compensated by a favourable perturbation pressure gradient induced by gravity waves, which acts on top of the fixed geostrophic pressure gradient.
    b. Line 561: “The wind turbines extract approximately 70% of the total energy input W_kef+W_gpg,wt.” I have a problem with this statement as part of the energy might come from flow deceleration inside the wind farm (i.e., the advective term). By decelerating the flow releases kinetic energy that can be used by the wind turbines or that can be dissipated. The total energy input needs to take this advective source term into account. As shown by Allaerts and Meyers (2017, J. Fluid Mech.), the combination of advective sources and perturbation pressure can be up to 15% of the energy input.
    c. Line 601: “As this influx is proportional to the BL height, much more energy is available for the wind turbines in the case CBL-1400-7D than in the case CBL-700-7D. This results in a higher kinetic energy flux …” I have difficulty with accepting that more energy available automatically leads to higher energy

fluxes without actual data or equations to support this. It would for example be interesting to look at the kinetic energy content of the flow above the wind farm and see how that decreases.

d. Line 612: "This redistribution is done by a favourable perturbation pressure gradient inside the wind farms and reaches power densities of approximately 1 W/m2 (not shown in Fig. 10)." Energy redistribution due to pressure gradients in the case of wind farm blockage is a major topic nowadays, so showing the actual perturbation pressure gradient contribution could be really helpful in this ongoing debate. I see no reason to exclude it from the figure.

Other scientific questions:

- Abstract and result sections: When talking about power density for very large wind farms, the work by Enrico Antonini comes to mind (see, e.g., Antonini and Caldeira, 2021, papers in Proc. Nat. Sci. Acad. and Applied Energy). How does your work compare to their estimate of the power density limits for large wind farms?
- Abstract and result sections: the clockwise flow deflection above the boundary layer is stated as one of the major findings occuring above very large wind farms and not for small wind farms. However, as mentioned on line 464, the effect might be overestimated due to conservation of mass flux in the FA. Moreover, the sensitivity to the Rayleigh damping layer is said to be out-of-scope, so there is no way of knowing how important this effect is, if it occurs at all. This seems quite a controversial result to me, so I wouldn't highlight it without additional sensitivity analysis or explicitly mentioning the uncertainty about this effect.
- Line 196: "While steady-state turbulence is reached after only a few hours, achieving a steady-state mean flow can take several days." How did you assess stationarity of turbulence and mean-flow quantities? Which quantities did you track and when did you consider it to be steady state (e.g., time rate of change less than 1% per hour)?
- Figure 7 and 8: It is interesting to see that the IL returns to its original height for CBL cases but not for the NBL cases. Is this due to faster wind farm wake recovery in the CBL cases, allowing horizontal convergence?
- Figure 8: x = 120km is identified as the near wake. Could you say how far downstream that is from the wind farm trailing edge, i.e., what is the x-position of the last turbine at this y position?
- Figure 8: The wind speed excess in the FA for NBL-700-5D is greatest at 120km and smaller (but not zero) at 180km. I'm surprised to see that the wind direction at 180km is more ageostrophic than at 120km. This means that there should be some other force at play other than the force balance between pressure gradient and Coriolis force. Or is the flow not in a geostrophic balance due to some kind of inertial oscillation in space (almost like a Rossby wave I suppose, a spatial wave supported by the Coriolis effect?) Can you elaborate? Same goes for NBL-700-7D and SBL case.
- Figure 8: For some cases, the vertical axis includes part of the Rayleigh damping layer. This is particularly visible for the neutral cases. The vertical profiles show non-physical behaviour in the Rayleigh damping layer. I think it would be useful to remind the reader of that by adding a line or a gray zone to indicate the location of the damping layer, or by only showing results below the damping layer.

- Line 505: Why not obtain the reference power for a single turbine using the same inflow profiles of the corresponding case, rather than using the neutral case as a reference for all? Now the reference power does not account for differences in wind shear or wind veer under non-neutral stability conditions, so you are attributing effects due to stability at individual turbine level to wind farm efficiency.
- Line 591: "Figures 10c and d show that a doubling of the BL height has approximately no effect on the energy input by the pressure gradient on the undisturbed inflow." I'd say there is a clear difference: W_gpg,BL is 50% higher in CBL-1400-7D than in CBL-700-7D. Are you maybe referring to the sum of W_gpg,BL and W_gpg,wt? Please clarify.
- Line 629: "Some tuning of the domain height and the boundary conditions was necessary to capture this phenomenon correctly." How did you assess whether you capture gravity waves correctly?
- Line 640: "…, the turbine spacing of very large wind farms should be at least 7 rotor diameters to achieve an acceptable wind farm efficiency." What value of wind farm efficiency do you consider acceptable? Isn't this up to the wind farm developer? Similar statements were made in the abstract, so please adjust accordingly.

Minor/technical comments:

- Line 44: For your information, the Lillgrund wind farm has also been simulated with LES by several authors. Consider adding references to, e.g., Churchfield et al. (2012, *AIAA*) and Nilsson et al. (2015, *Wind Energy*) to make this list more complete.
- Line 60: I'm curious as to how much computational resources (in terms of core hours, e.g.) were needed for simulations of this scale. Could you comment on this in the paper?
- Equation 2: The buoyancy term makes use of angular brackets, but the use of these is not defined in the text. Do they indicate horizontal averaging? How do you do that in the main domain which is horizontally not homogeneous?
- Equation 3: The transport equation for internal energy is stated in terms of the potential temperature, while the buoyancy term in equation 2 uses virtual potential temperature. How did you relate these two quantities? Note that neither of these quantities/symbols are defined in the text. Please define all symbols in the text.
- Line 78: You mention that overbars are used to indicate filtered quantities, but the LES equations 1-3 contain no overbars for the velocity components nor for the potential temperature. The notation is inconsistent. Either add overbars consistently in the equation, or mention explicitly that overbars are not shown for these parameters (so only shown for the subgrid-scale stress and heat flux) and that u_i etc. correspond to the filtered quantities.
- Line 142: Please describe what you mean with radiation boundary conditions. Does this simply mean zero velocity gradients and a fixed pressure value, or do you impose more intricate boundary conditions?
- Section 2.4: What boundary condition did you set for the potential temperature at the top of the domain? Please specify in the text.

- Line 225: Basically, subsidence velocity is chosen such that w_sub * Gamma = dtheta_0/dt. Maybe add the equation to clarify how w_sub is set.
- Line 241: "…, as this is the *correct* surface forcing method for SBLs." The word correct sounds quite strong and implies that applying a surface heat flux is incorrect. Rather, Basu et al. show that applying a surface heat flux can only represent weakly stable conditions and the surface forcing method is preferred. I wouldn't go as far as saying that there is a *correct* method for applying boundary conditions for SBLs.
- Section 2.5: This section doesn't add much, so I would try to integrate it elsewhere. Particularly the discussion on how to calculate resolved turbulent fluxes is, as far as I know, standard practice in LES, so I don't think it needs to be mentioned here. I'm not familiar with the term *temporal eddy-correlation method*, is that something you introduced yourself? If not, add references to where this term is used as well.
- Line 335: "Because the wind farms in this study also have a finite size also …" The word "also" appears twice in this sentence, I think one of them can be removed.
- Caption of figure 5: "The forces are normalized … and are horizontally averaged over one turbine spacing along x and y." The main text speaks about horizontal averaging over 4 spacing in x and 2 in y. Please clarify.
- Line 361: Could you state by how much percent the streamwise geostrophic pressure gradient force has increased? It is hard to see in the figure and a quantitative value would be instructive.
- Line 406: "The convective velocity scale is greater in the case CBL-1400-7D than in case CBL-700-7D …" It would be useful to mention the definition of the convective velocity scale here (or add the actual value to table 1) to show why it differs between these two cases.
- Equation 13 and fig 10: Use consistent naming: W_f or W_vkef. There is also a typo on line 561 W_kef.
- Line 637: "The achieved power density of turbines in the upstream part if the wind farms …" This should be "… *of* the wind farms …"