Response to reviewer

We would like to thank the reviewer for their constructive comments. We responded in detail to them below in blue.

General
The manuscript addresses the crucial consideration of Wind Farm Parameterization (WFP) in Numerical Weather Prediction (NWP) models, focusing on power production forecasting and the local weather impact of wind farms. The authors integrate Explicit Wake Parameterization (EWP) into the widely used HARMONIE-AROME model, comparing it with the existing FITCH wind farm model. The EWP relies on the actuator disc thrust force and incorporates wake vertical expansion within grid resolution limitations, without additional Turbulent Kinetic Energy (TKE) like the Fitch Scheme. The study rigorously compares these WFPs in HARMONIE-AROME and WRF across different settings and measurements. The paper is well-written, presenting information concisely, offering practical insights for research and application in wind energy.

Main point
A notable drawback of EWP, compared to FITCH, is the absence of the TKE source term. The paper argues for this by assuming that the heterogeneous part is a component of the mean flow; thus, an additional TKE source is not necessary. However, I contend that for turbine wakes, especially considering rotational motion and tip vortex variations that are subgrid scale and cannot be resolved by the mean flow in the mesoscale model. Furthermore, the TKE may arise from vertical shear due to the high vertical resolution; however, horizontal shear also cannot be resolved with a resolution of a few kilometers. The lack of TKE consideration affects wind profiles and wake recovery, leading to underestimation of wake effects above hub height. A more in-depth exploration of this limitation in the discussion would enhance guidance for WFP selection.

We fully agree that the correct parameterization of the TKE effect of wind turbines is important. In EWP wind turbines are only an implicit source of TKE due to the formation of TKE due to vertical shear in the wake wind profile. As you pointed out correctly, this study hints that an explicit TKE source might be required to fully account for the effects of the turbines. To make that clearer, we added the following sentences to the conclusion section:
"Nevertheless, this study indicates that taking only the implicit TKE formation due to vertical shear into account is not sufficient. Instead, an explicit source of TKE is required to consider the TKE formation from the rotational motion of the rotor as well as from tip vortices. Furthermore, this study showed that EWP also exhibits a different wake recovery at hub height as well as a different vertical wake profile of TKE, wind speed and other parameters. The reason for these differences are both the vertical wake expansion considered in EWP as well as the missing explicit TKE source. However, observations of the vertical profile in the wake were not available for comparison and thus further studies are necessary to investigate the correct shape of the profile in the wake."

Other Points:
1. Equations (2-4), the core of EWP, are complex. The derivation process should be either shown or referenced to aid in understanding the underlying physical assumptions. Furthermore, it needs clarification whether these additions are new or identical to the EWP in WRF.

Thank you for pointing this out. We have added the following at the beginning of section 2.1 "The theory and derivation of EWP are described in detail in Volker et al. 2015. For convenience we repeat the main concepts here. "

In addition, we added in the end of section 2.1 "This implementation is similar to the implementation of EWP in WRF except that turbulent diffusion coefficient, K, is different, since it is provided by a different PBL scheme. Having EWP implemented in HARMONIE allows for comparisons of different WFPs in HARMONIE as well as comparisons of WFE in HARMONIE and WRF."

2. Figure 4 raises questions about Ct being larger than 1 at low ambient wind speed. An explanation in the manuscript would enhance understanding.

Figure 4 shows Ct for the complete operating conditions, i.e. at very low wind speeds below 3 or 4 m/s depending on the turbine type the turbine is not operating and therefore no Ct value defined. Thus, the thrust coefficient is always smaller or equal to 1 and for low ambient wind speeds above cut-in Ct approaches 1. Hence, Ct is not larger than 1 at low ambient wind speed. To make this clearer, we have added "for the complete range of operating wind speeds" in "The thrust and power curves of these three wind turbine models for the complete range of operating wind speeds are shown in Fig. 4. Note that Ct is always smaller or equal to 1, since below 3 ms^{-1} or 4 ms^{-1}, depending on the turbine model, the turbines are not operating and therefore Ct is not defined. The Ct curves cannot be extrapolated to lower wind speeds."

3. Line 197-197: "the standard deviation of errors (STDE) assesses the non-systematic error". I am not sure how useful of STDE in this paper (mentioned only once later in the paper.) or in general. For example, consider a two cases with equally large bias, where the first one has a correlation of 0 (i.e. random error+systematic error), and the second one has a correlation of 1 (totally systematic). However, STDE for the two cases can be equal, which limits the intepretation of it.

We agree with you that for the two cases that you describe the STDE could be the same. Thus, we agree that the STDE standalone is not a good metric to describe the accuracy of the model, since the interpretation is limited in that case. However, in combination with other metrics such as the BIAS and the correlation, the STDE completes the picture of the model evaluation: the BIAS assesses a systematic deviation of the simulation from the observation while the correlation coefficient assesses the strength of a linear relationship between the simulation and the observation. The STDE is similar to the correlation, but is not dimensionless and can therefore give additional insight. Thus, we think that the STDE along with the other error measures provides a useful addition.

In the manuscript, we use the STDE along with BIAS, RMSE and correlation to assess the overall simulation quality. All metrics are given in Tables 3, 5 and 6 and are therefore mentioned more than once in the manuscript, albeit not in the text. We therefore think that the STDE is a useful addition to the other metrics.

We have added the following to the manuscript to make it clearer: "CORR and STDE evaluate similar aspects of the model performance, but STDE is not dimensionless and therefore gives additional insights." and "While two different models can perform the same in terms of one error measure, e.g. the same correlation coefficient, they might perform differently in terms of another error measure, e.g. different BIASes. Thus, having four different error measures has the advantage that different aspects of the performance can be evaluated."

4. Line 113: The abbreviation "IFS" needs clarification.
We have changed the sentence containing IFS to: "We run HARMONIE in forecasting mode using hourly boundary fields from the Integrated Forecasting System (IFS) global model at ECMWF as lateral boundary conditions."

We also clarified the abbreviation NEA: "We simulate the Northern Europe DMI domain A (NEA)".

5. Figs. 8, 11, 13: I don't understand why the authors show two line for each simulation instead of one time-interpolated line to the measement time. The figures are also quite small to see. Some suggestion: rearrange the right legend to the top or bottom and elimitate the white space in each subplots; change color codes in to a more conistent way (e.g. red for HAR, blue for WRF, solid for FITCH, dashed for EWP, dotted for NWF).

We show two lines for each simulations, since the model output time step does not correspond to the exact transect time of the aircraft, since the model output time step is 10 minutes for WRF and 15 minutes for HARMONIE. In addition, the passage of the transect took the aircraft around 10 minutes during which the conditions might have changed. This is also highlighted by the two lines. For clarity we added to the manuscript: "For the models, the two closest model output time steps (Table 1) to the transect times are shown to highlight the temporal variability of the conditions during the passing of the aircraft over the transect. Thus, for a good match between model and observations, the observations should be within the shaded area of the model output. Overall the wind speed at 250~m matches well with the observations for some transects, even though WRF overestimated the 10~m wind speed if compared to SAR (Fig. 7)."

We moved the legend to the bottom to make the figures wider. However, the long caption of Figure 8 restricts the height of the figure and thus it cannot be enlarged further. However, the online version of the manuscript enables to zoom in and investigate dome details.

Thank you for your suggestion for a different color scheme scheme. However, we feel that we are already using a consistent color- and line-style coding with brighter colours and more dense broken lines for WRF and darker colours and more loose broken lines for HARMONIE, as described in the caption of Figure 8: "WRF simulations (brighter colours, densely broken lines) and HARMONIE simulations (darker colours, loosely broken lines)". EWP is always shown by dotted lines, FITCH by dashed lines and NWF by solid or dashed-dotted lines for WRF and HARMONIE, respectively. Having only two colours and the same line-style would make it more difficult to distinguish the model outputs, especially for color-blind people.

We modified Figure 8, 11 and 13 as well as A1 and A2 to make the figures consistent.