General Comments:

The article by van den Broek et al. uses a free-vortex model with three different wind turbine representations, 2 different actuator discs and one "actuator line", to study wind turbine wakes behind yawing turbines. First, a numerical sensitivity study is performed followed by a comparison against three wind tunnel data sets with increasing complexity and dynamics.

The written is generally fine and the article holds interesting aspects, although they are mainly related to the dynamical experiments from other studies, where one of the three studies is currently unpublished. However, the main problem is that the validation is misleading to a certain degree, as a number of important phenomena are neglected. Additionally, there are issues with the associated accuracy/error quantification, which is not up to state-of-the-art in terms of validation practices in general and for vortex methods in particular, where the proposed model lacks novelty. Therefore, I unfortunately can not recommend the article for publication.

Specific Comments:

1. Validation.

Validation studies are important to evaluate and continue to improve numerical models and build confidence and the scientific foundation for any subsequent studies. So to some degree it is very difficult to achieve a "conclusive" validation (line 11), as there will be need of more detailed validations in the future, even if this study was published. But more importantly, I don't think the present study fully qualifies as a validation study for several reasons.

a. The modeled physics are not the same as in the measurements, and the authors neglects the impact rather casually. There are three main aspects of lacking physics:
    i. no shear modeling
    ii. uniform loaded turbines (it is unclear if the experimental turbines have uniform loading distributions, but most wind tunnel turbine models will not)
    iii. no turbulence modeling

    All, but particular the last two, are known to be very important for the wake development. No turbulence is not equivalent to 1% turbulence intensity and the authors does not address the limitations related to their uniform load modeling. Below are several references, which address previous work on both uniform loading and other tools based on vortex methods, which captures the three aspects.
b. The numerical verification/sensitivity study in 3.1-3.4 is generally fine, and important to test. But I am not convinced that these tests are sufficient to prove that the methods are good enough to solve the dynamic yaw problem, nor that the chosen metrics are representative, when the different model fidelities clearly affect the

physics. For instance, the wake expansion is presumably different and therefore the rotor averaged deficit (Equation (20)) could be misleading and potentially cause the deviation seen in Figure 9.

c. The balance of computational and physical considerations seem off. Computational costs are important, but so are the physics. There are at least 5 instances in the article (line 192-195, 211, 228, 370, 407), where simplifying choices are made purely based on computational efficiency without any consideration of the physical impact. It does not provide a fair comparison to first make choices based only on computational efficiency, and then analyse/conclude that the method most sensitive to these numerical choices (ALM) is performing poorly (line 374-376 and 380-384). The conclusion that ADM provides the best balance of computational cost and accuracy therefore appears selective as the choices inevitably lead to those conclusions. Furthermore, this balance is not quantified, and the article would benefit from reporting the computational cost of the three different turbine representations. Before, seeing the computational cost, I would still question if the recommended method is actually fast enough for dynamic wind farm control applications. Finally, if computational speed is the only criteria, why does the authors not simply use the well-known FLORIS model (https://github.com/NREL/floris)?

2. Accuracy and acceptable errors.

The article does not sufficiently deal with and analyze the required accuracy. For example, line 398 states that added complexity is not necessary for control purposes. I would argue that the wake behind a yawed turbine is more complex than the wake behind a non-yawed turbine, so the modeling choices should reflect that, see Boorsma et al. reference below. Again, if the complexity is not important, then why not use FLORIS? I suggest the authors to compare the VFW results with FLORIS, which have previously been shown to provide similar accuracy at low costs, particular for mid- and far wakes. The improvements of power (Figure 11) are also only assessed qualitatively to be of "similar magnitude". How does this compare to the quantified VAF and NMAE? The NMAE and VAF does indicate qualitatively that most of the dynamics are accounted for (line 418), but the remaining 6-9% could cause the consistent under-prediction of ADM seen in Figure 12, which is presumably related to lack of turbulence.

Finally and most importantly, I think the authors should reflect more on what accuracy is required and hence what are acceptable errors for wind farm control studies. Many previous studies on wind farm control have suffered from large/unknown errors and uncertainty, so I think the authors need to address if errors of 6-9% are acceptable relative to the potential power increase seen in wind farm control studies. Therefore, it is also too casual to state that the accuracy in operating point is more important than exactness in predicted power (line 445-446). How does the authors intend to determine “optimal operating point” with methods that give inaccurate power predictions? The understanding and reduction of errors and uncertainties are critical to make wind farm flow control successful and applicable on industrial scale, as discussed in the large international benchmark reported here:

- Göçmen, T., Campagnolo, F., Duc, T., Eguinoa, I., Andersen, S. J., Petrović, V., Imširović, L., Braunbehrens, R., Liew, J., Baungaard, M., van der Laan, M. P., Qian, G., Aparicio-Sanchez, M., González-Lope, R., Dighe, V. V., Becker, M., van den Broek, M. J., van Wingerden, J.-W., Stock, A., Cole, M., Ruisi, R., Bossanyi, E., Requate, N., Strnad, S., Schmidt, J., Vollmer, L., Sood, I., and Meyers, J.: FarmConners wind farm flow control benchmark – Part 1: Blind test results, Wind Energ. Sci., 7, 1791–1825, https://doi.org/10.5194/wes-7-1791-2022, 2022.

To be honest, I would not trust future results on wind farm control based on the presented validation of the VFW tool. Therefore, I think the aim of validating a simple model to use for wind farm control studies is counterproductive from a scientific (and societal) perspective, where we need more accuracy and confidence in results, not just speed.

3. Lack of novelty and references.
Vortex methods have a lot of potential, and a number of tools based on current state-of-the-art vortex methods exists, which have been validated in more detail than the present work. The novelty of the present work is therefore mainly related to the dynamic experimental results published elsewhere, not the model. Several of these tools include turbulence modeling, and should be properly referenced. A couple of recent publications are provided here, but I suggest to search more on these tools and current state-of-the-art vortex methods:

- Bergua, R., Robertson, A., Jonkman, J., Branlard, E., Fontanella, A., Belloli, M., Schito, P., Zasso, A., Persico, G., Sanvito, A., Amet, E., Brun, C., Campaña-Alonso, G., Martín-San-Román, R., Cai, R., Cai, J., Qian, Q., Maoshi, W., Beardsell, A., Pirrung, G., Ramos-García, N., Shi, W., Fu, J., Corniglion, R., Lovera, A., Galván, J., Nygaard, T. A., dos Santos, C. R., Gilbert, P., Joulin, P.-A., Blondel, F., Frickel, E., Chen, P., Hu, Z., Boisard, R., Yilmazlar, K., Croce, A., Harnois, V., Zhang, L., Li, Y., Aristondo, A., Mendikoa Alonso, I., Mancini, S., Boorsma, K., Savenije, F., Marten, D., Soto-Valle, R., Schulz, C. W., Netzband, S., Bianchini, A., Papi, F., Cioni, S., Trubat, P., Alarcon, D., Molins, C., Cormier, M., Brüker, K., Lutz, T., Xiao, Q., Deng, Z., Haudin, F., and Goveas, A.: OC6 project Phase III: validation of the aerodynamic loading on a wind turbine rotor undergoing large motion caused by a floating support structure, Wind Energ. Sci., 8, 465–485, https://doi.org/10.5194/wes-8-465-2023, 2023.
- S. Perez-Becker, J. Saverin, R. Behrens de Luna, C. Combreau, M.-L. Ducasse, D. Marten, A. Bianchini, 2022, "D2.2. Validation Report of QBlade-Ocean", FLOATECH Deliverable, Technical Report
- Ramos-García et al. 2023, "Multi-fidelity vortex simulations of rotor flows: Validation against detailed wake measurements", https://www.sciencedirect.com/science/article/pii/S0045793023000154
- Alvarex, E. J. and Ning, A., 2022, "Reviving the Vortex Particle Method: A Stable Formulation for Meshless Large Eddy Simulation", https://doi.org/10.48550/arXiv.2206.03658

A similar study have also been performed against wind tunnel data using LES, where the uniform and non-uniformly loaded ADM were compared, and found that uniformly loaded ADM is not sufficient to capture the correct physics. The authors should reflect on such results:

- Lin, M. and Porté-Agel, F.: Large-eddy simulation of a wind-turbine array subjected to active yaw control, Wind Energ. Sci., 7, 2215–2230, https://doi.org/10.5194/wes-7-2215-2022, 2022.

Other studies show that even actuator lines are not necessarily accurate enough for turbines operating in high yaw with the increased complexity of full scale experiments, and this is particular related to the simplifying assumption of using 2D airfoils, see for instance the following reference:

- Boorsma, K., Schepers, G., Aagard Madsen, H., Pirrung, G., Sørensen, N., Bangga, G., Imiela, M., Grinderslev, C., Meyer Forsting, A., Shen, W. Z., Croce, A., Cacciola, S., Schaffarczyk, A. P., Lobo, B., Blondel, F., Gilbert, P., Boisard, R., Höning, L., Greco, L., Testa, C., Branlard, E., Jonkman, J., and Vijayakumar, G.: Progress in the validation of rotor aerodynamic codes using field data, Wind Energ. Sci., 8, 211–230, https://doi.org/10.5194/wes-8-211-2023, 2023.

Technical Corrections:

1. Unpublished third study

The third experimental study in 4.1.3 is not yet submitted for publication so it is hard to understand the experimental setup. Perhaps it is the lack of drawing showing the experimental setup, but it is hard to understand how moving the second turbine laterally out of the wake of the first turbine should correspond to a change of wind direction?

2. Language and minor corrections

A number of wordings are in my opinion misleading. Here are a number of major ones, but the article would improve by general clarification of specific comments.

a. The FVW is in my opinion not a surrogate (line 5 + line 59). A surrogate is typically data-driven model based purely on regression of input-output without physical modelling. The FVW does solve simplified physical equations. Have a look at: https://en.wikipedia.org/wiki/Surrogate_model
b. The three different models (ADM, ADMR, ALM) are representations of the turbine, not wake models (line 73). The FVW models the wake based on the forcing from the different turbine representations.
c. "Actuator line" is the method used to represent individual blades in CFD (for instance LES) where forces are smeared numerically, while the term used in vortex methods is typically "lifting line", see your own reference van Kuik, 2018.
d. The wake transitions (line 231) does not seem "natural", but as a purely numerical artifact.

e. The trends of the model results might be meaningful, but it does not mean that they are accurate (line 458).
f. Equation (12) does not show pressure jump, but vorticity. They are of course related as stated in 136-137, but rephrase to match the equation.
g. line 161: To my understanding, the bound vorticity of the root vortex should be opposite to the tip vortices.
h. Line 342: If you have an induction factor of a = 0.33, then CT = 0.89 with same number of significant digits.