Review of the paper "Field comparison of load-based wind turbine wake tracking with a scanning lidar reference" by Onnen et al.

The paper deals with an investigation into the accuracy of a Kalman filter-based wake center tracking strategy using real field data. The topic is already addressed in literature but now the Authors were able to compare the tracking outputs with a reference, which is expected to be more accurate, i.e. a scanning Lidar.

The topic is worth investigating and the work well executed. The manuscript is clear and well-organized. I recommend publishing it. At the same time, I suggested some minor corrections and comments. Among all, I consider those listed under the section "Important comments" as something that, if addressed, may significantly improve the manuscript.

Important comments:

1. Page 6, equations (8): much is written in relation to the cut-out frequency of the low-pass filter that models the wake meandering, but very little is said about the adequacy of the model. Can Authors enlarge the treatment, possibly including a reference? Moreover, is it possible to detail how the Authors considered the variability of mean wind speed u_∞ and turbulence intensity in such a model?
2. Page 9, equation (9): Please, notice that transforming blade loads through the Coleman transformation yields two pieces of information (M_tilt and M_yaw) really close to the nodding and yawing moments that are easier to measure (e.g. strain-gauges on main bearing). Surely, they are not identical (e.g. the nodding moment there will be biased due to rotor weight) but they should carry the very same informative content requested by the detector. Given the fact that "The rotor azimuth angle information of WT2 was not available" (see line 74), this consideration could be practically relevant rather than a pure mathematical comment. Please comment.
3. Line 170: "The yaw and tilt moment depend on the wake position"; this is true, but they depend on other parameters, such as the shear layer magnitude. Authors cope with this by adding the terms b and c in eq. (11), to model, among all, also the impact of shear. However, the shear is variable too. Can Authors comment on this fact?
4. Line 210 and subsequent: important considerations. Good to see them here. Can the Authors provide insight into the possible application of the methodology using field data where one cannot control and decide a priori the inflow conditions to use to train the model?

Minor comments:

1. Line 30 and 64: missing references.
2. Figure 9: consider increasing plot dimensions.

3. Figure 9b: consider the possibility to add a new figure, representing the error between “Geometry” and “Geometry + Jimenez” versus the lidar estimate. This could improve the interpretation of the results.
4. Section 3.1.2: at what downstream distance is the speed deficit measured by the Lidar?
5. Comments on Fig.12: it is important to notice that the estimator is able to detect wake impingement on both sides of the rotor (left/right). I totally understand that maybe Authors considered it self-evident or trivial, but this is the very first capability that a wake detector must have.