Dear **reviewer #1**,

Thank you for the second round of reviewing.

In Figure 1, we have now combined close-up photos of the kite and the ground station during flight operation in the exact configuration that was used for the instrumented flight test introduced in Section 2. The flying kite in Figure 1(a) now complements the kite in launch position in Figure 2 and the CAD geometry of the kite in Figure 3. The three configurations are decisively different. Compared to the CAD geometry, the aerodynamically loaded membrane wing shows substantial deformations, especially noticeable here in the billowing of the canopy, which inevitably leads to a contraction of the trailing edge geometry (which is not visible). We think that for the introduction, this is a good first figure, giving an overview of the entire system, now in a lot more visual detail than before, and also including the ground station. That, considering that many readers will not be familiar with the technology. Figure 2 then visually outlines the experimental setup and sensor positions, but with the kite on the ground, and unloaded. Figure 3 introduces the bridle line layout, geometrical dimensions, reference frames, and terminology.

Dear **reviewer #3**,

Thank you for the second round of reviewing. We were aware of the points you addressed, but still searching to strike the balance between the level of detail of the explanations and readability. Thank you for your helpful suggestions.

Overall assessment:

This is my second time reading this work. Overall the description of the study and the model has been greatly improved and I am much more satisfied reading it. So, I don't have further comments on this section of the paper. Concerning the aim of this paper, (understanding how the mass of the KCU effects the turning and the development of models to account for this), the authors have definitely made some important progress in this direction, so this is a good paper. However, there is one point that I am uncomfortable with, on line 471 within the conclusions the authors state that the model resolves pitch and allows AOA computation. However, the results show large discrepancies when comparing numerical pitch and measured pitch, so I think that authors cannot make this point in the conclusions without further context or qualification (see notes below). Furthermore, I still think that the discussion around the source of errors can be further improved with 1 or two sentences at different points in the text. I encourage a more broad discussion on the source of errors at specific points in the text, but at the same time I can see a broader discussion in other places of the paper. So these criticisms might be more opinionated in nature, so I leave it up to the authors to decide whether these suggestions would improve the paper.

Important concerns on the conclusions

Starting at line 471: The authors state that resolving pitch allows for accurate AOA. However, the comparisons given in the paper are unsatisfactory. So I feel that it cannot be said in the conclusions that models lead to improved predictions for AOA, I think some further qualification and context is needed for this. Like the fact that 2 dof model could in theory give better AOA results, by better resolving pitch, however in this study there are still large differences and further study is needed to understand the sources of these errors. I think the authors need to be more precise to avoid (unintentionally) misleading the reader.

Response: We added some qualification to two paragraphs of the conclusion for which we differentiated between accuracy of solving the pitching motion and adequacy for determining the angle of attack:
".. since the developed model aims for simplicity to increase computational efficiency, it does not incorporate all relevant mechanical effects, such as tether elasticity. In addition to solving the motion dynamically, it could be necessary to refine the configuration of the kite model in order to increase the accuracy of solving the pitching motion and explain the observed differences between the measured and computed pitch."
and later:
"Further study is needed to assess how refined the pitching motion needs to be solved to accurately calculate the angle of attack of the wing."

Additional comments on the discussion (i.e. my more opinionated comments)

Line 342: Some speculation is given on the source of differences in the tether force between models and measurements. The hypothesis given, is that they are due to errors in the reconstruction. Another could also be that sharp variations of the tether force are attenuated by deformation and damping effects that are not considered in the model. As an outsider to this study, it's not my place to give my own speculation ... but as an overall, a question in this study is to what extent are errors caused by shortcomings in the modelling, problems in the measurements or analysis of data? I would be curious to hear the authors opinions on additional sources of error.

Response: We expanded the explanation to make it more complete. Text added: "The wing control input being a source of error is affirmed by coinciding, unexpected tether length results computed with the steady-rotation states. Consequently, also the tether acceleration control input will be a source of error since it is derived from these results. Errors introduced by model deficiencies, such as neglecting kite deformation and elasticity and damping of the tether, are expected to be overshadowed by relatively high errors due to the input."

Starting at Line 375: There is a discussion on the differences between the pitch measurements and the predicted pitch in the models. There are significant differences in the turns. One factor is that in the model, the KCU bridal and wing assembly are treated as a rigid body in the model, the pitch is the orientation of that rigid body. Where as the measurements are from point locations on a wing that is undergoing deformation (this point on deformation is clear by the measurements). So I see a couple alternative explanations for these large differences. 1) the measurements are not sufficient to get an average pitch estimate of the whole assembly due to local deformation effects 2) the modelling fidelity is insufficient to get accurate predictions (i.e. wing/bridal deformation is important). Again, the authors hypothesize that the differences are attributed to the reconstruction errors ... Again it's not my place to speculate on what's really happening, I am just a reviewer ... but the point is that in the absence of supporting data, it would be better for future study to have a broader discussion on the different sources of error and limitations of both the model and the experiment. If the reader wanted to repeat this study and collect improved results, is it really just better reconstruction that they should focus on? Personally I am skeptical on this point ... but even on that point, you could run a sensitivity study on the reconstruction results to assess that point (i.e. do your results change significantly with small variations of reconstruction). I am not suggesting that such a study is carried out for this paper ... It's just that I am seeing speculation that isn't backed up by data, without consideration of alternate sources of error and the fact that analysis that could be used to back-up these assertions missing. So I feel the explanation of these errors could be strengthened.

Response: To be more complete, we addressed the different sources of error and qualification statements: ".. Contrastingly, the two models exhibit systematic differences during the turns, with the differences in pitch being particularly substantial. This can be attributed to substantial transient effects during the turns, which are disregarded by the steady-rotation states. Although the dynamic result lies closer to the average measured pitch during the turns, it does not exhibit a similar peak. This discrepancy is expected to be caused by multiple sources of error. First, the actual wing motion that is causing the peak in pitch during the turns might not be accurately reconstructed in the flight trajectory reconstruction. This is affirmed by the large imposed modifications due to the relatively high uncertainty of the position measurement during the turns. Second, the pitch of the kite assembly calculated with the rigidly-linked two-point kite model is inadequate to describe the desired pitching behaviour of the kite during this highly dynamic manoeuvre. Consequently, a higher-fidelity model might be needed to obtain a suitable, higher-resolution pitching motion description. And third, assuming that the kite model is adequate, the available measurements could be insufficient to estimate the pitch of the kite assembly due to the measurement of local wing deformation."

and

".. Despite including transient effects, the dynamic model does not necessarily show a better agreement with measurements than the steady-rotation-state model. This suggests that improving the method for solving the motion may not be effective unless the configuration of the kite model itself is refined to compute pitching motion more accurately. However, a definite conclusion cannot be drawn because the uncertainty of the measurements might distort the view of the model validity."

To debate the rigid-kite assumption more, Fig. 14 is added showing the deformations of the kite together with a paragraph in the "Kite attitude validation" section.

Section 4.4 deals largely with the same pitch problems that I have already highlighted from section 4.3. From my reading of figure 14, I feel the greatest error in pitch occurs during the turning manoeuvres in real-out, yet in the reel-in I feel that there is better agreement. This leads me to think that there is something occurring in the reel-out turns that is not adequately captured. I am not sure if these full cycles results can help explain this ...? If you agree I am curious if the authors could say something about this in the paper?

Response: we are not convinced that the pitch is more accurately solved during reel-in than during the straight sections in the reel-out phase. It is primarily the highly dynamic turns that give a poor agreement with the measurements. Therefore, we kept Sect. 4.4 as is.

<u>Minor errors</u>

Line 355: "The available measurements *are* useful ..."

Response: reformulated