Review of the paper entitled "Model updating of a wind turbine blade finite element beam model with invertible neural networks", by Pablo Noever-Castelos et al.

General Comment

The paper deals with an updating procedure for complex wind turbine blade models. The methodology uses modal data (frequencies and shapes) to estimate the properties of the model (Young's and shear moduli, densities, etc…) through invertible neural networks. It seems that the work is totally built on a technique developed by the Author in a previous publication (Pablo Noever-Castelos, *Model updating of wind turbine blade cross sections with invertible neural networks*, submitted to Wind Energy). The paper is well-written and clear enough (I believe that a researcher unfamiliar with neural network should be able to get the main points of the work). The presented results generally support the goodness of the proposed methodology, even if they may be better commented.

There are three main concerns that should be addressed to provide a paper which is worth publishing, related to 1) link between present and previous publication of the Authors, 2) the collinearity (ambiguities) among features and 3) interpretation of the obtained results in light of the Sobol indices. These three concerns are better explained in "Major comments". This document also lists some "minor comments", that I hope the Author will accommodate in a revised version of the manuscript.

My final recommendation is to accept the manuscript only if the three "Major comments" will be adequately addressed.

Major comments

1. The link between the previous paper of the same Authors (*Model updating of wind turbine blade cross sections with invertible neural networks*) and the present one should be better highlighted, and the differences clearly stated. As far as I have understood, in the previous paper, only the final sectional properties are considered, while here the blade model is more complex and comprises detailed sectional descriptions. Moreover, the method for replacing the sensitivity computations (see sec. 4.5) is new. If so, the Authors should also comment on the need of such an addition of complexity. What is the eventual balance between a higher complexity in the model (and in the updating process) and the performance potentially achievable? Does it worth it? Finally, probably the word "beam" in the title indicates a simpler blade model with respect to that used in the work which is characterized by a fully three-dimensional description. Please, check.
2. In Section 2.3 "Selection of Sobol Indices", the Authors use the Sobol Indices to select only those parameters which significantly affect the system response (see lines 150-151 "we aim to consider only features which have a significant impact during at least one event at one location, thus containing enough information for the updating process.". This is totally correct, but it is certainly possible that combinations of input features may lead to similar outputs. This implies that one cannot comprehend 'who does what', and in turn cannot generate a robust model. During the identification of physics-based models, this problem is called collinearity (ambiguity, in this paper) and is handled by looking at the sensitivity matrix (e.g., SVD of sensitivity matrix, Cramèr-Rao bounds). Even if neural networks are employed, the collinearity problem should stay the same, because it refers to the intrinsic properties of the system. Since the Sobol indices are just metrics to quantify the sensitivity of outputs respect inputs, I imagine that collinearity problem (if present) may be found looking at such indices as well. The Author should comment and possibly extend the treatment.

3. Interpretation of the results in Section 4.1, (see especially fig. 9): At this point, it is essential to create the link between the goodness of the prediction and the Sobol indices. This will ease the comprehension of the results. For example, feature33 has a high Sobol index but an accuracy rather poor (R^2=0.8, with a significant spread around the regression line). On the other side, feature4 has a very low Sobol index (0.11 close to the selected threshold) but has an excellent accuracy. Why? I imagine that the link between the sensitivity analysis and the estimation accuracy should be stronger than what we see in such results. Please, comment and possibly explain thoroughly the obtained results.

Minor comments

- Line 85: Please, correct “resimulationn”
- Line 95: Please, remove comma in “Sobol derived, the 1st order Sobol index…”
- Line 113: symbol “N_{FE}” appears here for the first time, but its meaning was not previously defined.
- Line 131: “All applied variances are approximately twice the permitted manufacturing tolerances”. Probably with the word “Variances” the Authors refers to the difference imposed to the parameters to perform the sensitivity analysis. If so, the word “variances” could be misleading as it often indicates a statistic metric.
- Line 141: does the sentence “… and the six degrees of freedom of each finite element beam node NFE are saved and…” refer to modal shapes?
- Tab. 2: It would be interesting to plot Sobol indices as function of the blade span, parameterized with respect to the typology of the element, so as to give an idea on how the observability changes as function of the blade span. The Author may try convert the table into a plot.
- Section 3: the description of the network could be improved. In particular, it could be important
    - To clarify what is new with respect to previous works; it seems that the network is totally built on previous activities, and no dedicated updates were conducted for the present research.
    - To clarify the reason why this network type is better suited to the application at hand. From this point of view, I would expect here a connection with the Introduction, and especially with the three points listed in Sec. 1.2. Why is the present network able to handle more complex problem than those already studied in literature? How can the present network evaluate uncertainty in the results? Why is the present network able to create a generalize model not focused on a particular condition?
- Line 231: check spelling of “resimulationn”.
- Section 4.1: The analysis is good and interesting, but here comes again the main question: as in a major comment, is it possible that the poor accuracy of some parameters may be connected to collinearity problems?
- Section 4.1, fig. 9: Instead of using in “Feat_x”, the reader could benefit from subtitles with the physical meaning. So, he/she does not have to jump to Tab.2
- Section 4.1, fig. 9: At this point, it is essential to create the link between the goodness of the prediction and the Sobol indices. This will ease the comprehension of the results. For example, feature33 has a high Sobol index but an accuracy rather poor (R^2=0.8, with a significant spread around the regression line). On the other side, feature4 has a very low Sobol index (0.11 close to the selected threshold) but has an excellent accuracy. Why? I imagine that the link between the sensitivity analysis and the estimation accuracy should be stronger than what we see in such results.

- Lines 316-322: These lines and the previous section talk about something that I had in mind since the beginning of the manuscript: the different properties of each section (Young modulus, densities, etc.…) may contribute together to the final sectional stiffness, and eventually it is hard to distinguish among those properties looking at global pieces of information (modal data). This, however, refers to an intrinsic problem of the systems. When Authors write "we can state that the cINN should correctly predict the total mass and the stiffness contributions in a global manner …", at least for me, they report something rather obvious. Please, comment and, if needed, clarify.
- Fig. 15: The analysis underlying this plot is interesting. I was wandering whether a similar conclusion can be derived from the Sobol analysis of Sec. 2.1. I guess that features belonging to blade root and tip be associated to both lower estimation accuracy (see Fig 15) and lower Sobol indices (from Tab. 2). Please, verify and comment.
- Line 386: "cINN correctly captures the global model behavior with respect to mass and stiffness distribution.". What about the blade center of gravity position, which is a value simple to be measured? This data can be used in the estimation process. Was it done?
- Section conclusion:
    - The sentence "invertible neural networks are highly capable to efficiently dealing even with an extensive wind turbine blade model updating" should be better explained. In fact, the estimation problem is solved but still the updating process results accurate only for global model characteristics (see line 386: "the cINN correctly captures the global model behavior with respect to mass and stiffness distribution"). I suggest stressing this fact.
    - Lines 461-462: "The ambiguities are captured very accurately by the network.". What do the Authors mean with this sentence? Does it mean that the cINN is able to get rid of ambiguities and not-identifiable combinations of features and perform the estimation accurately for the rest of the features? If so, maybe the sentence should be clarified.