General comments:

The manuscript "The importance of process interactions and parameter sensitivity for modelling the carbon dynamics in a natural peatland" describes a calibration and sensitivity analysis of the CoupModel, using several variables measured on an eddy covariance test site in Sweden. Several variables, which describe carbon, energy and water fluxes, are used for calibration and a sensitivity analysis.

Overall the manuscript is not well written and difficult to follow. In wide parts corrections by a native speaker is required. However, the study is interesting and contains relevant aspects. Unfortunately, the actual presentation of the study is not convincing. Several parts are too fuzzy and too general, while other parts are too detailed. The objective is not clear and the conclusion does not provide any new information despite general knowledge about this field. I am really puzzled to rate this manuscript, as there are many concerns in almost all parts of the manuscript. However, as I see also the potential of the study, I rate it acceptable with **major revisions**, but I have to be clear, that only answering the comments below won't be enough to get the publication to an acceptable form. As there were too many issues, it was not possible to comment all in detail, but I tried to explain my concerns on some parts in more detail.

Overall comments:

The objective is not really clear. Reading the paper it seems like that all variables are needed to improve the quality of calibration, several parameters are interacting and more measurements are needed. I do not need a study to come to this conclusion. I miss out more numbers, that rate the quality, and real values, like how much quality do I miss out, if I calibrate only on one variable rather than on all variables (for R2, ME and NSE). There is no discussion about transferability of the results and the robustness of the results. Which results can be used in general and which are related to the CoupModel. I do not see the list of other studies as a discussion of transferability. If the authors want to include these studies, there need to be an analysis of the differences for the different approaches used for the different processes.

I am also not sure if the picked indicators describing the goodness of fit are well picked. The mean error will compensate strong negative and positive disagreements in the overall value, which do not reflect the quality of the model performance. I would like to see the root mean square error used instead. Also R2 is not a good value for model performance as it might be sensitive to extreme values, if not all parts of the data range are represented equally.

The state of the art method for calibration is Bayesian calibration, which is not mentioned in this study. At least in the introduction and maybe in the discussion this needs to be mentioned and explained why the here used method is as good, better or worse than the Bayesian calibration and what are the advantages and disadvantages.

Using several variables for calibration of models or a sensitivity analysis the pareto optimization would be an appropriate multi-criteria approach to address the subjective judgement of the model performance. However, at least this technique and/or other approaches for multi-criteria optimizations should be mentioned and discussed.

Specific comments:

Comments on the title: First, the model used in this study is not mentioned in the title. Second, the title contains only the carbon dynamics, while most of the variables that are considered in the analysis are energy and water fluxes. The title needs to be re-formulated and more precise.

Comments on the objective:

- For point 1. the authors do not identify processes, but parameters and variables, which are most sensitive in the model to simulate the target variables
- Point 2. is not well formulated and it is difficult to understand the objective.
- I am not sure about point 4. Why do you test the usability of measurements? The measured data are usable. Additional, it is not true that you can detect the missing measurement variables. You can detect model sensitivity and required data for the used version of CoupModel. Other models might need other parameter sets and might show different sensitivities. Also, the authors work out improved model performance by adding more variables in the calibration process, it still rise the question, if additional measurements are need, the model approach that simulates the process needs to be improved or the calibration approach needs to be improved.
- In the objectives it is not mentioned what the authors are actually doing. The model is not mentioned and the four points are not linked to any land use, model or analysis approach.
- The sentences after point 4 do not contain any useful information about the actual study, but only general information what you can do with an outcome from a sensitivity analysis.

Comments on the method section:

Section 2.2: The gap-filling of the climate data is explained, but not the gap filling of the EC data. The model description is far too long, but leaves crucial aspects out at the same time. There is also lack the scientific terminology.

Page 4 line 20: EC is not defined. Please add this on line 15 the same page.

Page 4 lne 28-29: C uptake by the ecosystem from the atmosphere

Section 2.3.3 There is no need to give a general introduction into soil hydrology.

Section 2.3.4 There is no need for a general introduction into phenological models, but provide the key information: used phenological model, the model is based on temperature sum and day length, parameters and settings. Also the description of allocation of carbon in the plant is too long and not well formulated. Especially, the labelling of parameters in the model do not contribute to a better understanding of the study.

Section 2.3.5 The section is too long, in some parts the scientific terminology is not used, the description of the processes is too casual and essential information is missing (e.g. turnover rates of the pools, decomposition follows first order kinetics). Based on the description I am not sure, if the model considers really different temperature sensitivities for fungi and bacteria (which would surprise me) and where the data about community size are coming from. I assume the authors mean that the SOC module contains a parameter that controls the impact of temperature on the decomposition rates and this factor was calibrated and tested for fungi and bacteria dominated soils. For the here presented study this doesn't matter.

Section 2.4.3 A couple of problems with the NEE values could be sorted by using the correction approach by Papale et al., 2006 (Biogeosciences, 3, 571–583). This would enable to solve the problems with extreme day values and the peaks for the night periods. I also wondering, if the gap filling tool, develop by Reichstein and Falge, is used to fill gaps for NEE?

Comments on the result section

Page12 lines 11-14 I understand that the soil water content is an important variable, which is difficult to measure and to simulate. This is not new and as this is known, this should be a central part of a sensitivity analysis. I think it is not enough to ask for more measurements, which is always a good answer to all problems with simulations. First, I miss a discussion of the measurements of the soil water content, which is often done on a single spot rather than spatial distributed or in different depths. Second, there is no discussion of the footprint area of the EC measurements. If the footprint changes and the soil type or hydraulic properties differ on the test site, this might explain differences. Third, as the authors make a sensitivity analysis, it is possible to detect the most sensitive soil property and give at least the advice, which soil property should be measured to get better results with the CoupModel.

3.1 Parameter sensitivity:

I do not understand why the authors highlight the module dependency so strong. This analysis makes the study extremely model dependent. I think the authors

should relate the sensitivity to processes. I assume that the modules represent separate processes, but this is not necessarily the case.

Page13 lines 27-29 R2 and ME are contradicting in their goodness of fit: Is this an indication that these are not the best indicators to detect the quality of performance?

Section 3.4 Usefullness might be not a good word to describe the measured variables.

Comments on the discussion:

Wide parts of the discussion are not really a discussion, but do only compare qualitative findings of the study with other studies.

4.1 Parameter sensitivity

It is correct that the detection of sensitivity of parameters enable to concentrate the calibration on the main drivers, but how robust are the findings on this test site and how transferable are the results to other ecosystems or to other climate zones? Peatland in Northern Europe is a quite specific test site, so, is it possible to transfer the results to mineral soils? How transferable are the results to Central Europe or to the Mediterranean area? It is no problem, if the results are not transferable, but at least there need to be a discussion.

Page 19 lines 3-5: "While the existence of interactions between the processes and their parameters is supposed to be less dependent on site conditions and model structure, the exact shape of the connections as well as constraint parameter ranges might strongly depend on these factors. " This might be correct as the sensitivity analysis only represents effects of the model structure. However, by applying the analysis on a specific test site, the relevance of processes depends on the climate zone, ecosystem, land use, soil type, etc. This also effects the limitations for the data range of the considered parameters and variables. The relation and interaction might be different outside this range. Therefore, I wouldn't exclude the site conditions as relevant factors.

Page19 lines 14-16: It depends: Several models using the same approaches to describe processes. Therefore, the formulated hypothesis needs to be tested by compare the approaches used in the different models to be sure, that this correlations are really independent of the model structure.

Page 19 line 27 to page 20 line 2: I do not really understand how the implementation of open water bodies should explain the differences in the correlations. In the measurements H is more related to temperature and LE more to the water flows. Photosynthesis is the main driver for growth and photosynthesis is calculated by a light use efficiency function and, as written in this manuscript "….total plant growth is proportional to the net global radiation absorbed…..". Is it possible that the correlation of H and NEE can be explained by the calculation of photosynthesis by radiation, which is also the main driver for H, while LE is calculated in more complex equations with less direct correlation to radiation and temperature?

Page 20 lines 3-5: No, not necessarily. If you try to understand the pattern of data in advance, the used indicator for the goodness of fit can be picked sensible. E.g. there are variables with several values (e.g. night values) at zero or around zero. These values will have a strong impact on the ME as the models, usually, simulate the zero values during night quite well. The R2 can cope with the clouds around zero, but it is sensitive to single extreme values. Bottom line the used indicator for goodness of fit influences the outcome of the analysis and if the indicator is well picked, there are subjective judgements. Controversial results of different indicators need to be analysed to understand the reasons for the contradiction. Unfortunately, this analysis is missing in this manuscript.

Page 20 Lines 6 -15: Of course there are lot's of correlation between LAI and other variables, because these parameters use LAI. However, an analysis and discussion of the cited publications is missing. This would be a chance to bring the here presented study in the context of other studies. Instead of only mention the correlation, the authors could explain the different dependencies. E.g. I assume that LAI correlates with soil water content, if it is a dry, water limited ecosystem.

Page 20 line 17 temporal or spatial resolution? What means high resolution mm, cm or m; seconds, hours, days?

Page 23 Lines 10 – 15: I see the strong sensitivity of the soil hydraulic properties as relevant factor, but first, it is not that easy to measure these parameters and, second, I think the authors should provide an alternative method to derive better fits and quantify the reduction of quality by missing out soil hydraulic properties. An alternative method would be to calculate the soil hydraulic properties by pedo-transfer-functions (as mentioned in the model description). If do so, the sensitivity of single parameters (soil type, bulk density, field capacity (by itself) etc.) can be tested and it might be possible to get better calibration using this information or detect the most sensitive of these parameters.

Comments on the references

The publication of He et al. needs to be updated

Comments on figures:

- I would like to see a figure like Fig.5 also for actual values and not only for a prior and posterior comparison.
- The quality of the figures is not good

Comments on the supplement:

Table S1 I think there is no need to present parameter name in the model. I am even not sure if the module name provides any useful or needed information, but it might

be better to group the parameters instead (e.g. soil, hydrology, snow, vegetation/growth).

Table S2 is really needed, if you develop a model and publish it, but I do not see the use for the actual study. Most of the equations are standard approaches that are already described in the text.